Synthesis of *O*-Linked Glycopeptides Using Enzymatic Catalysis

A thesis submitted in partial fulfilment of the requirements for the degree of **Doctor of Philosophy** in **Biochemistry** at the **University of Canterbury**

David Lim

Supervisor: Prof. Antony Fairbanks

University of Canterbury

2015

Table of Contents

Abstract

Synthesis of *O*-Linked Glycopeptides Using Enzymatic Catalysis

Enzymes have many advantages over conventional synthetic methods. They can be used without the need of a protecting group strategy whilst maintaining high stereo- and regiochemistry. Over the past few decades, the use of enzymes has become commonplace because of their catalytic efficiency and their ability to be used under 'green conditions'. The use of glycosyl hydrolases (glycosidases), enzymes that catalyse the energetically favourable hydrolysis of glycosidic linkages, is also becoming increasingly widespread. In particular, the construction of glycosidic linkages using glycosidases has slowly gained prominence, however the tendency for the product of the reverse hydrolysis to be a substrate for the natural enzymatic hydrolysis reaction has plagued this process with low yields. To remedy this, glycosynthases, a class of mutant glycosidases, have been engineered to eliminate hydrolytic activity. In the presence of an activated donor substrate and an appropriate acceptor, glycosynthases are able to catalyze the formation of the glycosidic linkage and not the hydrolysis leading to higher product yields.

The work in this thesis focuses on the synthesis of activated *O*-glycan donors, and the expression, purification, and mutagenesis of glycosidase enzymes of the super-family GH101. Subsequent kinetic assays and glycosylation investigations with various acceptors using the wild type and mutant enzymes are described. Also, work on the synthesis of glycosides in water is presented.

Acknowledgements

I would like to sincerely thank my supervisor, Prof. Antony Fairbanks, for accepting me into his research group and letting me work on such an intriguing project. We may not have arrived where we wanted to, but the journey was certainly rewarding (for both of us, I hope). Thank you for the support and ideas you have given over the years as well as the good times outside the lab! I had a great time working and growing in your group!

Thanks also goes to Drs. Yusuke Tomabechi, Akshita Wason, and Andrew Watson. You all have taught me so much and inspired me to push further. Thank you all for your patience; to Yusuke, for endowing me the ability to consume beverages and stay standing for longer. To Akshita, for always being so very patient with me, especially when I made silly mistakes and wasted your time. And to Andrew, for always having nice quick solutions and being consistent with tea time.

My years doing research at Canterbury would not be as colourful, exciting, and entertaining without the members of the Fairbanks group of past and present. We grew from a group of two members after the Feb 2011 Christchurch earthquake to quite a decent 12 at its peak; Evan, Pragya, Govind, Vivek, Kajitha, Stewart (also for reading my thesis), Preeti, Sivasinthujah, Thomas, Xavier, Diane, Hélène, Aoife, and Maryne – thank you all the good and fun times we had together.

Further extension of my gratefulness goes toward the Parker group for letting me utilize their laboratory space and equipment, and especially Dmitri, Nicky, and Emma who

put aside a lot of their precious time to teach me a fair portion of what I know about practical biochemistry.

The technical staff at UC are a fantastic bunch! Thank you Drs. Marie, Amelia, and Alexander for your assistance in all things spectrometry and friendship too! You are all gifted in your areas and I have learned so much from you all. To Dr. Matt Polson, for literally being the all-rounder in the department – thank you for pouring your expertise and knowledge into the department. Thanks also extends to Wayne, Rob, Nick, Gill, and Laurie! You all make a fantastic team and the department is currently thriving because of it!

My gratitude also extends to Profs. Shinya Fushinobu and Makoto Ito from University of Tokyo, Japan and Kyushu University, Japan, respectively, for providing the plasmids for engBL and engEF for this project.

Thanks also goes to the Lord Rutherford Research Memorial Fellowship and the Marsden Fund for financially supporting me while I did my research.

A grand thank you goes to my family for putting up with me during my years in Christchurch, for providing me with good home to come back to every day. Thank you all for your patience and support.

And finally, I would like to thank my wife, Grace, who has endured such a long wait in Wellington for me to finish. For your unfailing support and encouragement, and for always believing in me. I love you!!

Abbreviations

The following abbreviations have been used in this thesis:

General:

δ	chemical shift
ν_{max}	wavenumber
^{o}C	degrees Celsius
µm	micrometer
µM	micromolar
Å	Angstrom
Ac	acetyl
ADMP	2-azido-1,3-dimethylimidazolium hexafluorophosphate
ampicillin^{r}	ampicillin-resistant
aq	aqueous
Ar	aromatic
at	apparent triplet (in NMR)
AU	absorbance units
BAIB	bis(acetoxy)iodobenzene
$BF_3.Et_2O$	boron trifluoride diethyl etherate
Bn	benzyl
br.	broad (in NMR)

Bz	benzoyl
c	concentration
C-$(His)_6$	*C*-terminal His-tagged
Calcd.	calculated
CEC	cation exchange chromatography
cm	centimeter
core 1	Galβ1,3GalNAc
core 2	Galβ1,3(GlcNAcβ1,6)GalNAc
core 3	GlcNAcβ1,3GalNAc
COSY	correlation spectroscopy
CuAAC	copper-catalysed azide-alkyne cycloaddition
d	doublet (in NMR)
D_2O	deuterated water
Da	Dalton(s)
DABCO	1,4-diazabicyclo[2.2.2]octane
DAST	diethylaminosulfur trifluoride
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DCM	dichloromethane
dd	doublet of doublets (in NMR)
DEPT	distortionless enhancement by polarization transfer

DFIH	2-fluoro-1,3-dimethylimidazolium hexafluorophosphate
DIPEA	di-isopropylethylamine
DMC	2-chloro-1,3-dimethylimidazolinium chloride
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DNP	2,4-dinitrophenol
E. coli	Escherichia coli
e.g.	*exempli gratia* (for example)
endo A	endo-β-*N*-acetylglucosaminidase from *Arthrobactor protophormiae*
endo M	endo-β-*N*-acetylglucosaminidase from *Mucor hiemalis*
engBL	endo-α-*N*-acetylgalactosaminidase from *Bifidobacterium longum*
engCP	endo-α-*N*-acetylgalactosaminidase from *Clostridium perfringens*
engEF	endo-α-*N*-acetylgalactosaminidase from *Enterococcus faecalis*
engSP	endo-α-*N*-acetylgalactosaminidase from *Streptococcus pyogenes*
Enz	enzyme
ESI	electrospray ionization
et al.	*et alia* (and others)
EtOAc	ethyl acetate
EtOH	ethanol

Fuc	fucose
g	gram(s)
Gal	D-galactose
GalNAc	*N*-acetyl-D-galactosamine
GlcNAc	*N*-acetyl-D-glucosamine
Glu	D-glucose
GNB	galacto-*N*-biose (Galβ1,3GalNAc)
h	hour(s)
HRMS	high resolution mass spectrometry
hLys	hydroxyllysine
HPLC	high performance liquid chromatography
HMBC	heteronuclear multiple bond correlation spectroscopy
HSQC	heteronuclear single quantum coherence spectroscopy
Hz	Hertz
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl β-D-thiogalactopyranoside
IR	infrared
J	coupling constant
kanamycin^r	kanamycin-resistance
k_{cat}	turnover number

kDa	kilo-Dalton(s)
K_M	Michaelis constant
L	litre
LB	lysogeny broth
lit.	literature
LNB	lacto-*N*-biose (Galβ1-4GlcNAc)
M	molar
m	multiplet (in NMR)
m	meta
m.p.	melting point
m/z	mass/charge ratio
M^+	molecular mass ion
Man	mannose
mbar	millibar
Me	methyl
MeCN	acetonitrile
MeOH	methanol
mg	milligram(s)
MHz	megahertz
min	minute(s)

mL	milliliter(s)
mM	millimolar
mmol	millimole
mol	mole
mol. sieves	molecular sieves
MS	mass spectrometry
MUC1	mucin 1
N-$(His)_6$	*N*-terminal his-tagged
N.D.	not detected
NeuAc	*N*-acetylneuraminic acid
nm	nanometer
nmol	nanomole
NMR	nuclear magnetic resonance
O/N	overnight
OD_{600}	optical density of a sample measured at wavelength of 600 nm
p	para
PCR	polymerase chain reaction
Ph	phenyl
Phth	phthaloyl
PMP	*para*-methoxyphenyl

PNGase F	peptide-*N*-glycosidase F
pNP	*para*-nitrophenyl
ppm	parts per million
Pra	propargyl glycine
p-TSA	*para*-toluenesulfonic acid
py	pyridine
q	quartet (in NMR)
quant.	quantitative
R	generic organic group, unless specified
R_f	retention factor
RMSD	root-mean-square deviation
RNase B	ribonuclease B
RP-HPLC	reverse phase high performance liquid chromatography
rpm	revolutions per minute
rt	room temperature
s	singlet (in NMR)
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
sec	second(s)
SEC	size exclusion chromatography
SGP	sialoglycopeptide

t	tertiary
t	triplet (in NMR)
TEA	triethylamine
TFA	trifluoroacetic acid
THF	tetrahydrofuran
t.l.c.	thin layer chromatography
TMSOTf	trimethylsilyl trifluoromethanesulfonate
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet
v/v	volume by volume
w/v	weight by volume
WT	wild-type

Amino acids:

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine

H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Declaration

I, David Lim, confirm that the work in this thesis represents my own, that the contribution of any supervisors and others to the research and to the report was consistent with normal supervisory practice, and that any external contributions to the research have been quoted and acknowledged by means of complete references.

Signed: 

Date: 18 November 2015

Chapter 1: Introduction

1.1 Significance of Carbohydrates

Carbohydrates are one of the four key classes of biological macromolecules, alongside proteins, DNA, and lipids.[1] The functional diversity of carbohydrates stands out because of their ability to form highly complex linkages and branched polymers. This complexity is made possible by several structural features of carbohydrates such as hydroxyl group stereochemistry, ring size, anomeric configuration, linkage position, branching position, reducing terminal attachment, and the linear sequence of core and linear branches. Such structural diversity means that a unit as small as a hexasaccharide is capable of forming 1.05×10^{12} possible carbohydrate structures, whilst a hexapeptide is capable of less than 1×10^{5} possible structures.[2]

1.2 Biological Importance of Carbohydrates

The understanding of the involvement of carbohydrates in a myriad of life-sustaining and life-threatening processes has seen carbohydrate chemistry evolve into a broad area of research that has captured the interest of the scientific community.[3] Initially, carbohydrates were thought to be simple molecules involved in the formation of structural materials, such as chitin and cellulose, and as a source of energy, such as glucose. More recently, carbohydrates have been shown to play vital roles in biological processes, leading to the development of the field, Glycobiology. The development of a complete understanding of the unique molecular properties of carbohydrates and their functions as bioorganic compounds is a remarkable challenge.

There are three main categories of protein-linked glycans: *N*-linked (where the sugar is attached to the amide group of Asn), *O*-linked (where the sugar is attached to the hydroxyl group of Ser, Thr, or hydroxyllysine (hLys)), and *C*-linked (where the sugar is attached to the carboxyl group of Trp).

Protein glycosylation is of vital importance, and is essential in many biological processes including cell-cell signaling,[4] microbial virulence,[5] inflammation,[6] and immune defense.[7] Post-translational modification of proteins by glycosylation can play a key role in protein folding,[8] and can also critically affect important protein properties such as conformation and stability,[9] susceptibility to proteases,[10] and circulatory lifetime.[11]

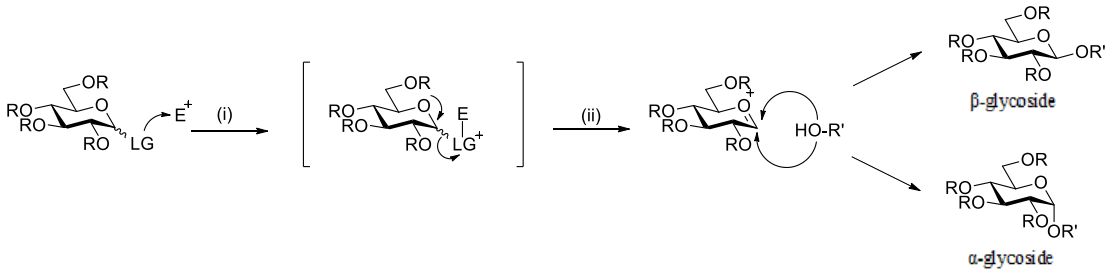
The study of the above effects, however, is made difficult due to the fact that glycoprotein biosynthesis and assembly is not under direct genetic control; this leads to glycoproteins being biosynthesized as complex heterogeneous mixtures. These glycoprotein mixtures are known as glycoforms, where different oligosaccharide structures are linked to the same peptide chain.

Due to their similar physical properties, these mixtures are usually inseparable. Interestingly, Rudd *et al*. successfully separated glycoforms of RNase B and demonstrated that different glycoforms conferred different activity to RNase B, and that heterogeneous mixtures of glycoforms provide a range of activities that could be influenced with fine-tuning of the glycan structure.[12] Therefore, to understand the function of individual glycoforms, methods to access homogenous glycoproteins is necessary.

1.3 Chemical Synthesis of Oligosaccharides

The chemical synthesis of complex carbohydrates has been a major focus of many research groups. The assembly of individual carbohydrate units is not a straightforward process. First, one must appreciate the complexity of the glycosylation reaction, which is capable of producing a wide range of products that vary in both stereo- and regiochemistry.

In a typical glycosylation reaction, an appropriate electrophilic promoter activates a glycosyl donor containing a leaving group at the anomeric centre. The resulting glycosyl cation is then reacted with the desired glycosyl acceptor to afford either an α- or β-glycoside (Scheme 1). For the synthesis of di- and oligosaccharides, this acceptor is a carbohydrate alcohol.



Scheme 1 - General glycosylation reaction. (i) Activation of glycosylation donor with an electrophile, E^+; (ii) capture of glycosyl cation with acceptor, R'OH to give either an α- or β-glycoside.

The choice of protecting groups and the sequence of their installation is very important for an efficient and successful synthetic route. This, in turn, may dictate the stereochemical outcome of the glycosylation reaction where the anomeric linkage can be formed in a 1,2-*cis* or 1,2-*trans* orientation with regard to the oxygen at C-2 of the donor and the new anomeric linkage. For example, an acyl group at C-2 (e.g. acetate, benzoate, phthalimide, or acetamide), can trap the glycosyl intermediate by means of

neighbouring group participation to form a cyclic oxonium ion (Figure 1). For glucosides, the α-face of the sugar is blocked by the neighbouring equatorial C-2 substituent, disfavouring attack by the acceptor from this direction, giving the β-product (Figure 1, top). For mannosides, the epimer of glucose at C-2, neighbouring group participation by the C-2 acyl functionality occurs on the top (equatorial) face of the anomeric centre, blocking acceptor access to this face and favouring the formation of α-mannoside products (Figure 1, bottom).

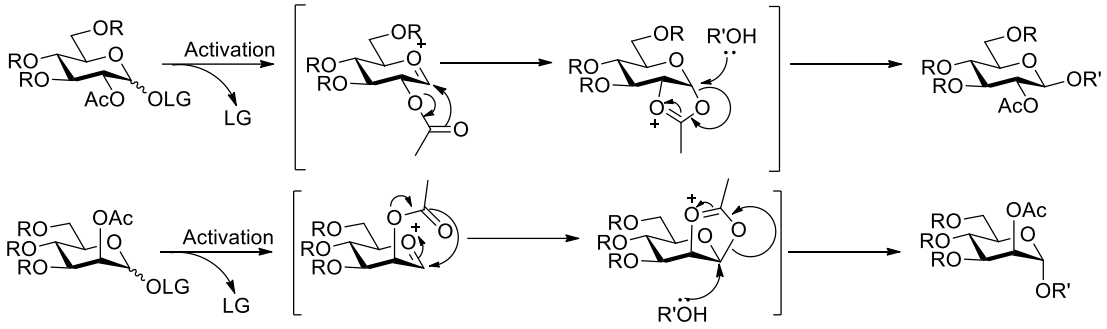


Figure 1- The formation of 1,2-*trans* glycosides by neighbouring group participation using glucosyl (top) and mannosyl donors (bottom).

Research and development over the past thirty years has been focused on discovering new anomeric leaving groups and promoter systems which give rise to high yielding, stereoselective, and efficient glycosylation reactions.[13–18]

1.4 *O*-Glycans

O-Glycans are one of two major types of protein-linked oligosaccharides, and result from the sequential enzymatic transfer of monosaccharide units to the hydroxyl groups of the side chains of several amino acids, mainly serine and threonine. *O*-Glycans are widely known for their roles as primary components of the intestinal mucus layer that covers the gastrointestinal epithelium. This mucosal layer forms an intestinal barrier

that protects epithelial and intestinal mucosal immune cells from potentially harmful luminal microflora and food components. The mucus layer is a dense, carbohydrate-rich matrix that consists primarily of mucins containing multiple serine and threonine residues modified by *O*-glycans, which account for 80-90% of the total mucin mass.

The roles of *O*-glycans are also functionally diverse. Generally, *O*-glycans are involved in roles associated with immunity, receptor-mediated signalling, regulation of enzyme activity, and protein expression and processing. Moreover, this modification has also been found to enhance protein stability, heat resistance, hydrophilic, and protease resistance to various proteins.[19,20,21] *O*-Glycans are also found in a wide variety of glycoproteins such as fetuin, human gonadotropins, glycophorin, and antifreeze glycoproteins.[22]

O-Linked glycosylation has also been suggested to play vital roles in determining the secondary, tertiary, and quaternary structure of a fully folded protein.[23] For example, the core 1 structures [Galβ1,3GalNAc-α-*AA*, where *AA* represents a serine or threonine residue] of the T1 fragment of glycophorin interacts with the peptide and limits the movement of the chain around a particular conformation which results in the stabilization of the conformation of the peptide backbone.[24]

Another form of *O*-linked glycosylation involves the attachment of *N*-acetylglucosamine (GlcNAc) to serine and threonine residues. Protein *O*-GlcNAcylation is regulated by two enzymes: 1) Uridine diphosphate *N*-acetylglucosamine:polypeptidyl transferase (OGT),[25,26] a glycosyl transferase that catalyses the transfer of GlcNAc from uridine diphosphate *N*-acetylglucosamine to

acceptor proteins and; 2) *O*-GlcNAcase (OGA)[27,28], a glycosyl hydrolase that catalyses the cleavage of *O*-GlcNAc from proteins. The cycling of *O*-GlcNAc on and off of proteins is catalysed by the reciprocal action of these two enzymes. The *O*-GlcNAc modification is known to regulate a diverse range of cellular processing including epigenetic regulation of gene expression[29], stress response[30], and circadian rhythm[31]. Decreases in brain *O*-GlcNAcylation has also been associated with Alzheimer disease.[32] More recently, *O*-GlcNAcylation has been found to be associated with the protection of nascent polypeptide chains from premature degradation by decreasing translational ubiquitylation.[33]

Unlike *N*-glycans, *O*-glycans are often classified on the basis of the sugar found at the reducing end. For mucin *O*-glycans, this monosaccharide is *N*-acetylgalactosamine (GalNAc) which is typically α-linked to serine or threonine. GalNAc is then extended by various glycosyl transferases, giving rise to up to ten different core structures (Figure 2).

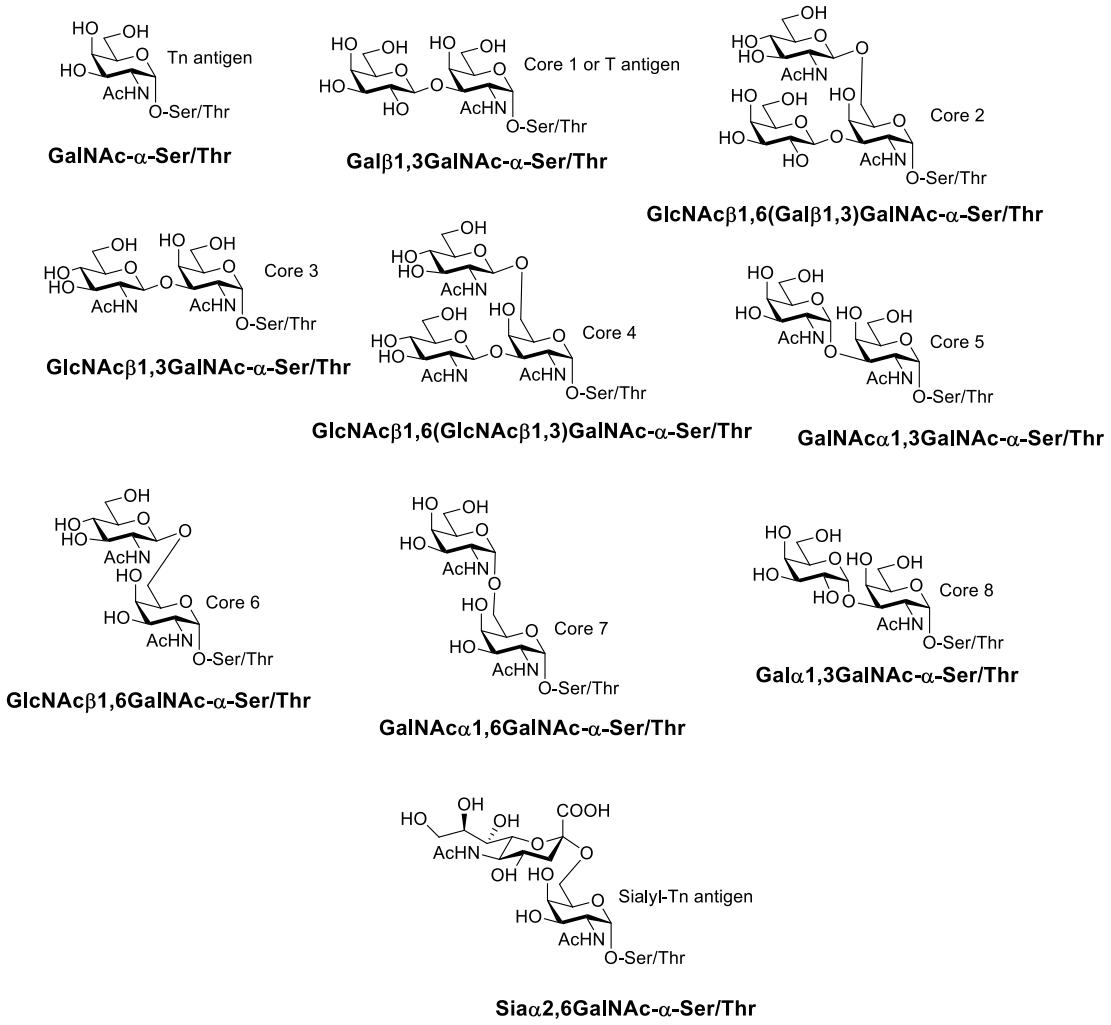


Figure 2 - Core structures of the *O*-Glycans

O-Glycans are assembled by the sequential addition of sugar units by several specific, membrane-bound glycosyl transferases in a highly controlled fashion.[19] The extension of the monosaccharide unit gives rise to important biological factors such as the blood group determinants (A, B, and H).

To date, there is no known consensus sequence for *O*-glycans, although there have been some noticeable patterns observed between the locations of *O*-glycosylation. For example, only serine and threonine residues that are exposed on the protein surface can be targeted by transferases for glycosylation.[34] Also, attachment sites associated with

O-glycosylation have been found in regions with high serine, threonine, and proline content.[35]

The structure of *O*-glycans present on different cell and protein surfaces has been extensively studied.[36,37] Traditionally, the first step towards structure elucidation of the glycan portion of a glycoprotein involves the glycan release from the polypeptide backbone for better analysis (although some non-liberating methods have been reported).[38] In the case of *N*-glycans, free reducing sugars can be conveniently and quantitatively released from glycoproteins by either the commercially available peptide *N*-glycosidase F (PNGase F) or a variety of endo-α-*N*-acetylglucosaminidases.[39]

In contrast, *O*-glycans are generally liberated by chemical methods due to the lack of 'universal' releasing enzymes. Reductive β-elimination (the Carlson degradation)[40], in which *O*-glycans are released from glycoproteins and then reduced to alditols *in situ* in alkaline sodium borohydride solution to prevent "peeling" degradation from occurring, is most commonly used (Figure 3). The disadvantage of this method is that the *O*-glycan alditols that are obtained cannot be further labelled with either a chromophore or fluorophore for spectrophotometric analyses.

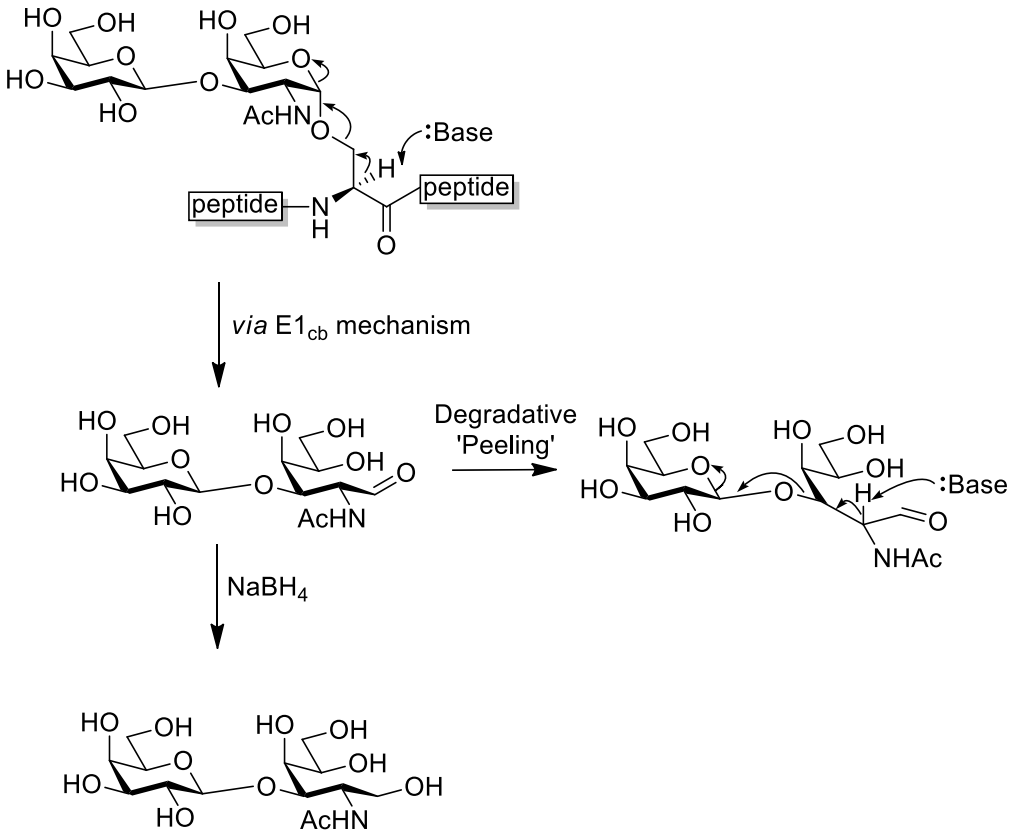


Figure 3 - Reaction mechanism for the release of *O*-glycans under basic conditions and the basis of the degradative 'peeling' reaction.

1.5 The biosynthesis of *O*-glycans

The biosynthesis of *O*-glycans is initiated by a family of enzymes, the Golgi-body localized polypeptide α-*N*-acetylgalactosaminyltransferases (ppGalNAcTs), which transfer a GalNAc unit from UDP-GalNAc to a Ser or Thr residue of a correctly folded protein, so beginning the first step in the biosynthesis of all mucin-type *O*-glycans.[41] This GalNAc-Ser/Thr unit, also known as the Tn antigen, can serve as an acceptor for at least three other Golgi glycosyltransferases: 1) core 1 β1,3-galactosyltransferase (core 1 β3GalT or T-synthase), which synthesizes the core 1 structure (also known as the T antigen); 2) core 3 β1,3-*N*-acetylglucosaminyltransferase (core 3 β3GlcNAcT), which synthesizes the core 3 structure and; 3) the sialyltransferase ST6GalNAc-I which synthesizes the sialyl-Tn antigen. Figure 4 outlines these biosynthetic pathways and

additional transferases involved in the extension of the core 1 and core 3 disaccharides to give complex structures.

1.6 *O*-Glycans in disease states and treatment

Unnatural changes to the extension of the innermost GalNAc unit have been associated with malignancies.[42] In 1969, terminal α-linked GalNAc were observed on human tumour cells. These cells bound to agglutinins from the snail-lectin HPA that recognizes the T antigen (Galβ1,3GalNAc).[43] Springer *et al*. then further reported that the Tn antigen was found at high levels in around 90% of breast carcinomas,[44] whereas little or no expression was observed in normal adult tissues. Further studies have associated the Tn antigen expression in cervical cancer,[45] lung adenocarcinomas,[46] colorectal carcinomas,[47] breast carcinomas,[48] and gastric carcinomas[49] with metastatic potential and poor prognosis.

These anomalies are associated with erroneous glycosylation in the biosynthetic pathways during the transport of proteins through the endoplasmic reticulum and Golgi bodies, which is often linked to a set of defective or absent enzymes. For example, a deficiency in UDP-GalNAc transferase 3 (polypeptide *N*-acetylgalactosaminyltransferase 3, GalNT3; EC 2.4.1.41), renders the organism unable to transfer UDP-GalNAc to Thr/Ser of a protein backbone. Patients with familial tumoral calcinosis (FTC), an autosomal recessive progressive metabolic disorder characterized by large calcium deposits in the skin and subcutaneous tissues, can have mutations in the *GALNT3* gene which codes for functional GalNT3.[50]

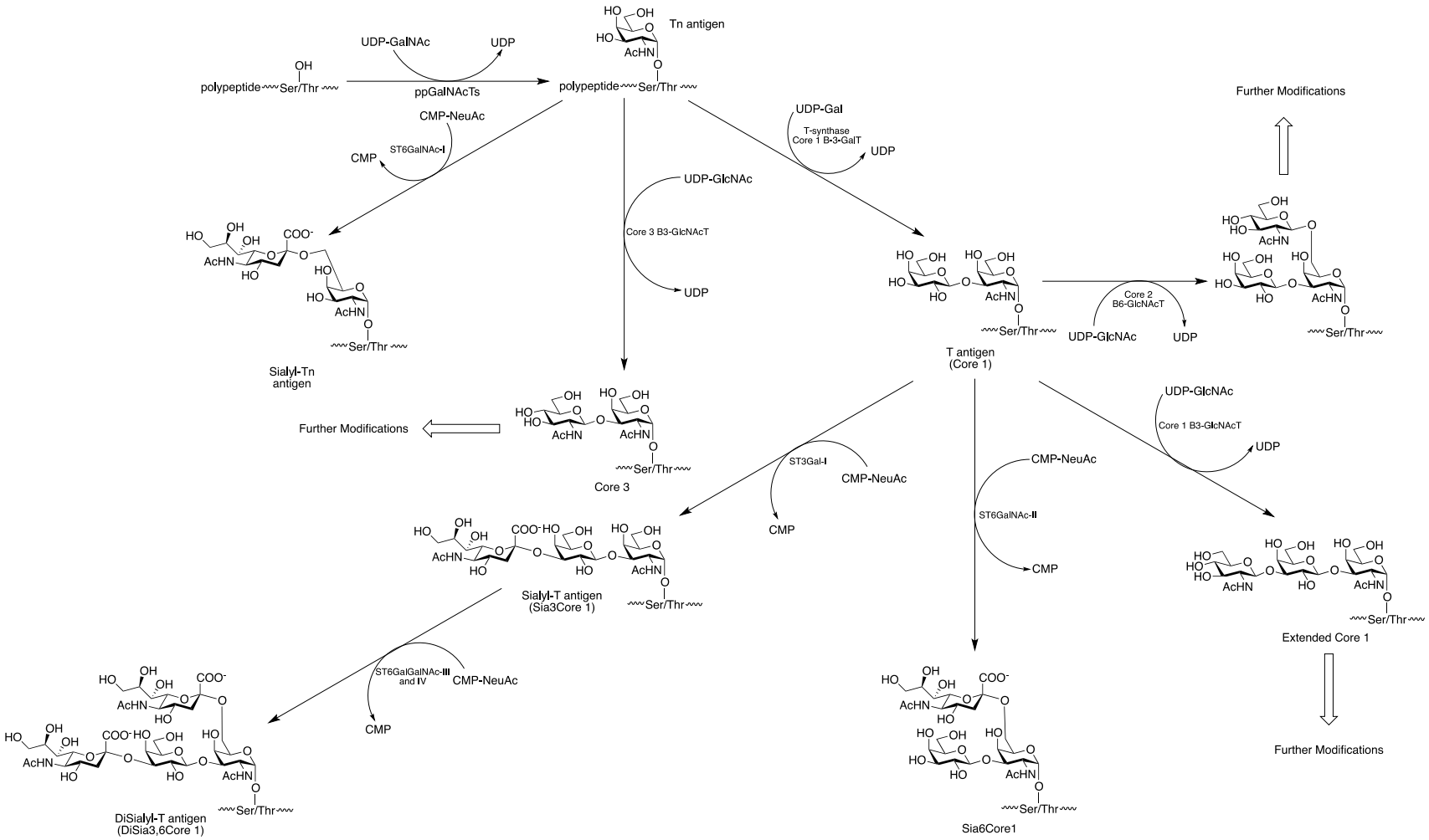


Figure 4 - The biosynthesis of the Tn antigen and different *O*-glycan core structures (adapted from Ju *et al*[56])

Changes in *O*-glycosylation of cell membrane glycoconjugates have been observed in neoplastic lesions from a variety of organs including lung, stomach, skin, and endometrium.[46] Wopereis *et al*. have published a comprehensive review on the many conditions associated with erroneous *O*-glycosylation.[51]

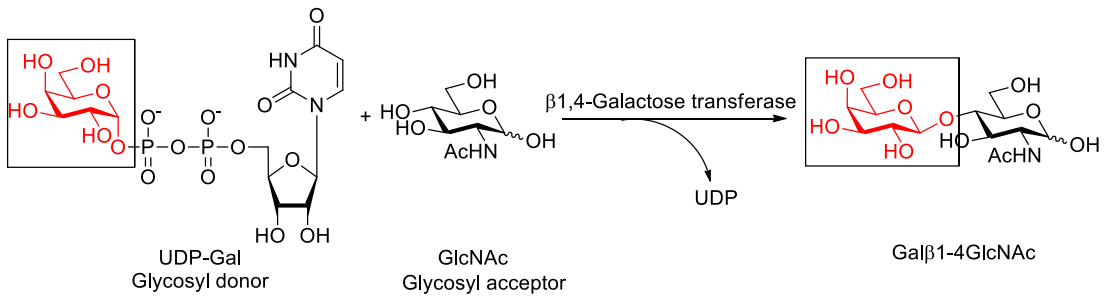
The T-antigen (Galβ1,3GalNAc-α-Ser/Thr) and its precursor, the Tn antigen (GalNAc-α-Ser/Thr), have been found to be over-expressed in more than 85% of human carcinomas, but not in healthy human tissues.[52–54] Since these truncated *O*-glycan variants are abundant on tumours, peptides containing defined *O*-glycan structures have been proposed as anti-cancer vaccines, which may generate an immune reaction against epitopes present on cancer cells.[55–57] Thus, identification of abnormal glycosylation patterns on cell-surface glycoproteins of malignant cells has motivated research efforts towards the development of cancer immunotherapy,[42,58,59] an alternative to established cancer therapies based on the possibility of activating the human immune system to recognize and effectively kill tumour cells.[60] The advantages of this approach over standard therapies, such as surgical removal of tumours, radiation, and chemotherapy, stem from the high selectivity and the delocalization of the immune reaction.[61]

Antitumoral vaccines are composed of tumour-associated carbohydrate antigens attached to peptides which provide stimulating epitopes for T cells.[62–64] Pioneering experiments were carried out by Springer *et al*, who used erythrocyte glycoproteins as antigens that were isolated from outdated, banked blood and treated with enzymes to expose the Tn antigen.[65] This worked based on skin hypersensitivity tests, which aimed to diagnose various carcinomas years before biopsies and X-ray examinations gave positive results. Encouraged by this research, recent attempts to generate vaccines have

been aimed at generating well-defined and pure Tn-antigen preparations. For example, Napoletano *et al.*[66], Kaiser *et al.*[67–69], and Boons *et al.*[70,71] have generated glycopeptide vaccines carrying clustered Tn antigens that were synthesised by either chemically or enzymatically transferring GalNAc to Ser/Thr residues of MUC1 peptide fragments. There has also been work published on the synthesis of non-natural mimics of the Tn antigen.[72]

1.7 Glycosyl transferases

Glycosyl transferases are responsible for the synthesis of glycans during glycoprotein biosynthesis. Glycosyltransferases generally act in a stepwise manner and are highly specific in their substrate requirements (Scheme 2).[73,74] There are two classes of glycosyl transferases: Leloir pathway glycosyltransferase have strict specificity for their respective nucleotide sugar donor and the acceptor, whereas non-Leloir pathway glycosyltransferases use sugar phosphates as the donor substrate.[75]



Scheme 2 – Glycan synthesis using glycosyl transferases.

In practice, glycosyltransferases have been very useful for the synthesis of oligosaccharides[76] with reactions often furnishing high yields. However, there are several difficulties associated with their usage for oligosaccharide construction. For example only a few enzymes are commercially available and complex sugar nucleotide

donors are required. Moreover, nucleoside diphosphate by-products are often inhibitors of the enzymes.[77]

More recently, a methodology called glycorandomization has slowly gained prominence, where non-native glycosyl 1-phosphates could be used as substrates for nucleotidyltransferase enzymes which had been modified to have increased promiscuity *via* directed evolution.[78] The resultant non-native sugar nucleotides have been shown to be effective glycosyl donors for glycosylation of a variety of natural products.[79–81]

1.8 Glycosyl hydrolases (glycosidases)

Glycosyl hydrolases are enzymes that catalyse the hydrolysis of the glycosidic linkage of glycosides, resulting in the formation of a sugar hemiacetal and the corresponding free aglycon. The importance of glycosyl hydrolases in biological systems cannot be underestimated as they play crucial roles ranging from degrading polysaccharide food sources through to manipulating glycoconjugates on the surfaces of proteins and cells. Due to the numerous glycosidic linkages found in biological systems, it is not surprising that there are correspondingly a vast number of glycosidases. To date, over 260,000 different glycosyl hydrolases have been identified and grouped into 135 families.[82]

There are two types of glycosidases: exoglycosidases, which cleave the glycosidic linkage to release the terminal residue of an oligosaccharide, and endoglycosidases, which cleave internal glycosidic bonds. Exoglycosidases are specific for a particular sugar and show varying specificity with regard to the particular linkage between two sugars. For example, whilst some sialidases will not discriminate between α2,3 and

α2,6-linked NeuAc, others are specific in that they only hydrolyse α2**,**3 linkages. Endoglycosidases are highly specific for their respective internal linkage of an oligosaccharide, for example, the endo-β-*N*-acetylglucosaminidases (ENGases) specifically cleave the GlcNAcβ1**,**4GlcNAc linkage of *N*-linked glycoproteins.

Although their primary role is to catalyse the hydrolysis of the glycosidic linkage, under controlled conditions, glycosidases can be used to catalyse the reverse reaction, that is, glycosidic bond formation.[77] As such, they have been used extensively as biocatalysts for oligosaccharide synthesis.

Alternatively, transglycosylases, a class of GH enzymes that can catalyse the intra- or intermolecular substitution of the anomeric position of a glycoside, can be used for the synthesis of glycoconjugates. In general, transglycosylases utilize the same mechanism as various retaining glycosidases whereby the reactive glycosyl-enzyme intermediate can be intercepted by water, or by another acceptor (such as a carbohydrate), to give a new glycoside. For example, the family GH70 comprises of transglucosylases produced by lactic acid bacteria. The enzymes in this family use sucrose as the D-glucopyranosyl donor to synthesise α-D-glucans with high molecular masses ($>10^6$ Da), releasing D-fructose as the by-product. In the presence of glucosyl acceptors, glycosylation occurs at the expense of polymer formation, and the competition between the two reactions is controlled by acceptor recognition by the enzyme. Moreover, polymer extension proceeds from the non-reducing end.

1.9 Mechanism of Glycosidases - Inverting and Retaining Glycosidases

Glycosidases are further sub-divided in two main classes by their mode of catalysis: inverting and retaining glycosidases.

Inverting glycosidases operate *via* a direct displacement of the glycoside by water (Figure 5a). As their name implies, the configuration of the anomeric centre in the product is inverted relative to that of the starting glycosyl substrate.

Retaining glycosidases utilize a double-displacement mechanism involving the formation of a glycosyl-enzyme intermediate (Figure 5b). This mechanism involves initial binding of the substrate to the enzyme followed by general acid-catalysed attack of a nucleophilic residue at the anomeric centre to form a glycosyl-enzyme intermediate. This intermediate is then hydrolysed by general base-catalysed attack of water at the anomeric centre, forming the product with retention of configuration and returning the enzyme to its original protonation state. Both the formation and hydrolysis of the glycosyl-enzyme intermediate likely proceed through transition states with substantial oxocarbenium ion character.[83–86]

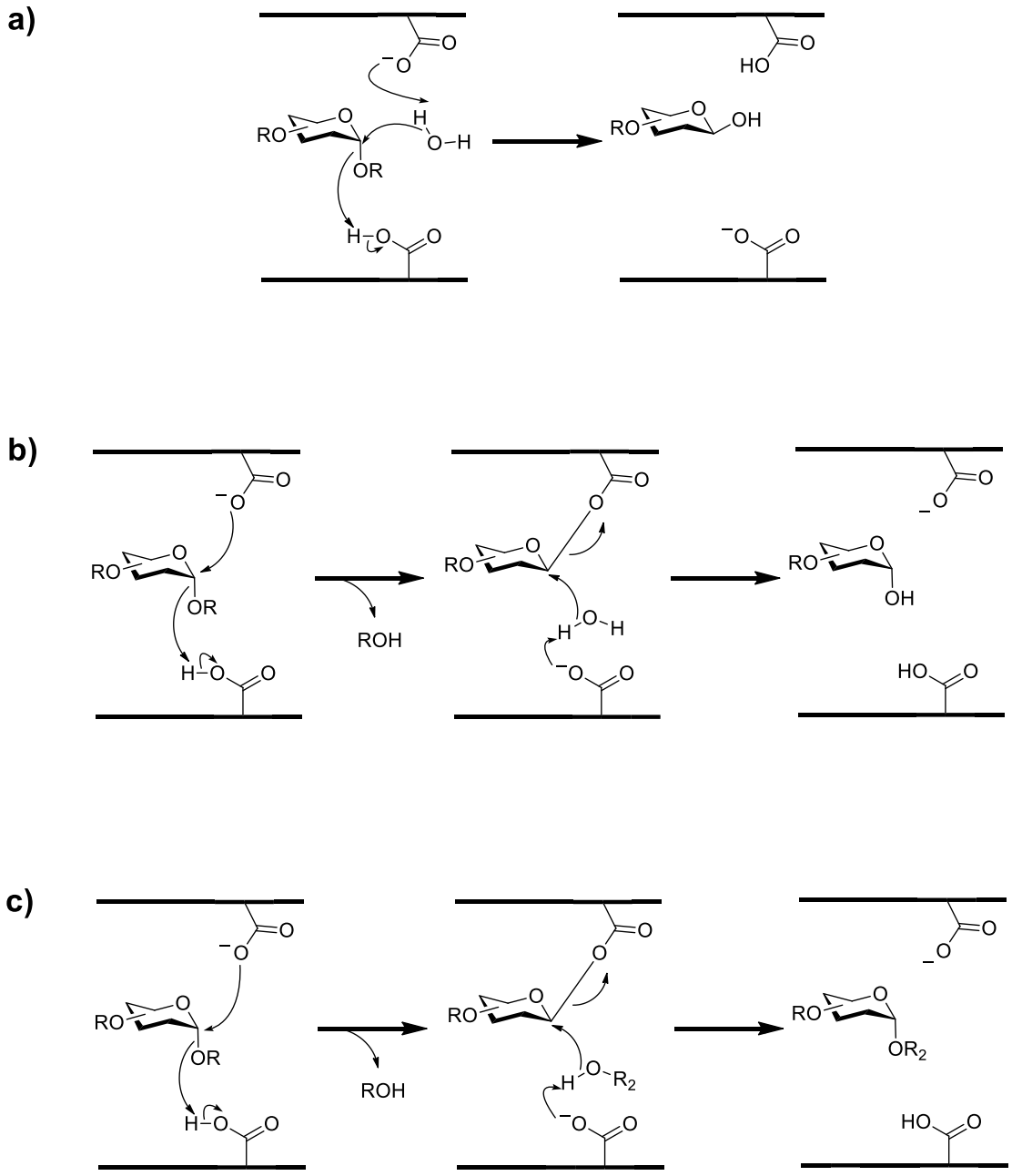


Figure 5 - Mechanism of glycosyl hydrolases. a) Hydrolysis catalysed by inverting glycosidases; b) Hydrolysis catalysed by retaining hydrolases; c) Transglycosylation catalysed by retaining glycosidases.

The transglycosylation reaction is considered a special type of hydrolysis wherein the glycosyl-enzyme intermediate can be reacted with other nucleophiles, such as alcohols or glycosyl acceptors (Figure 5c).

A major difference between the two classes is the distance between the two carboxylic acid residues: for retaining glycosidases, the distance is typically around 5.5 Å, whereas for inverting glycosidases it is usually around 10 Å.[87]

1.10 Uses of Glycosidases

Glycosidases have the primary function of cleaving glycosidic linkages, however they can catalyse the formation of glycosidic bonds if the conditions are correctly chosen. The transglycosylation activity of endoglycosidases has been proven useful for transferring oligosaccharides to various compounds, making them ideal for the synthesis of neoglycoconjugates.[88]

The two main approaches previously employed to enable the enzyme-catalysed formation of the glycosidic bond are thermodynamically controlled and kinetically controlled glycosylation.[77] The thermodynamic approach involves the reversal of the *in vivo* function of the glycosidase. This can be achieved by using large excesses of substrate,[77] organic co-solvents, high salt concentrations,[89] and/or high temperatures.[90] However, these methods still result in low yields and poor regioselectivity.

Alternatively, kinetic control relies on the formation of a reactive intermediate from an activated glycosyl donor that is subsequently reacted with a glycosyl acceptor to form a glycosidic bond more rapidly than it can be reacted with water. In such cases, the yield is optimised by utilising conditions where the rate of hydrolysis is slower than that of transglycosylation. Typically this can be achieved by using high acceptor concentrations and activated donor substrates. *p*-Nitrophenyl and fluoro-glycosides are

just two of many common activated donors that have been used in glycosidase-mediated transglycosylation reactions.[88,91,92]

1.11 Glycosynthases

To overcome the drawbacks of low yields of glycosidase-mediated transglycosylation reactions, glycosynthases, a new class of mutant glycosidases were introduced. The first glycosynthases were derived from retaining exo- and endo-β-glucosidases by replacing the active site nucleophile with a non-nucleophilic residue such as alanine.[91,93,94] Introduction of this residue resulted in a correctly folded enzyme, but which was catalytically inactive since it lacked the residue required to form the α-glycosyl-intermediate.

The importance of the mutation can be appreciated if one considers the mechanism of the retaining β-glycosidase (Figure 6a). As mentioned in Section 1.9, retaining glycosidases operate by a double-displacement mechanism involving the formation of a glycosyl-enzyme intermediate. General base-catalysed attack of an appropriate acceptor (water or alcohol) forms the product with retention of configuration. Removal of the catalytic nucleophile disrupted the natural mechanism of the enzyme (Figure 6b). Without the ability to form the glycosyl-enzyme intermediate, the glycosidase could not catalyse the hydrolysis of its natural substrates.

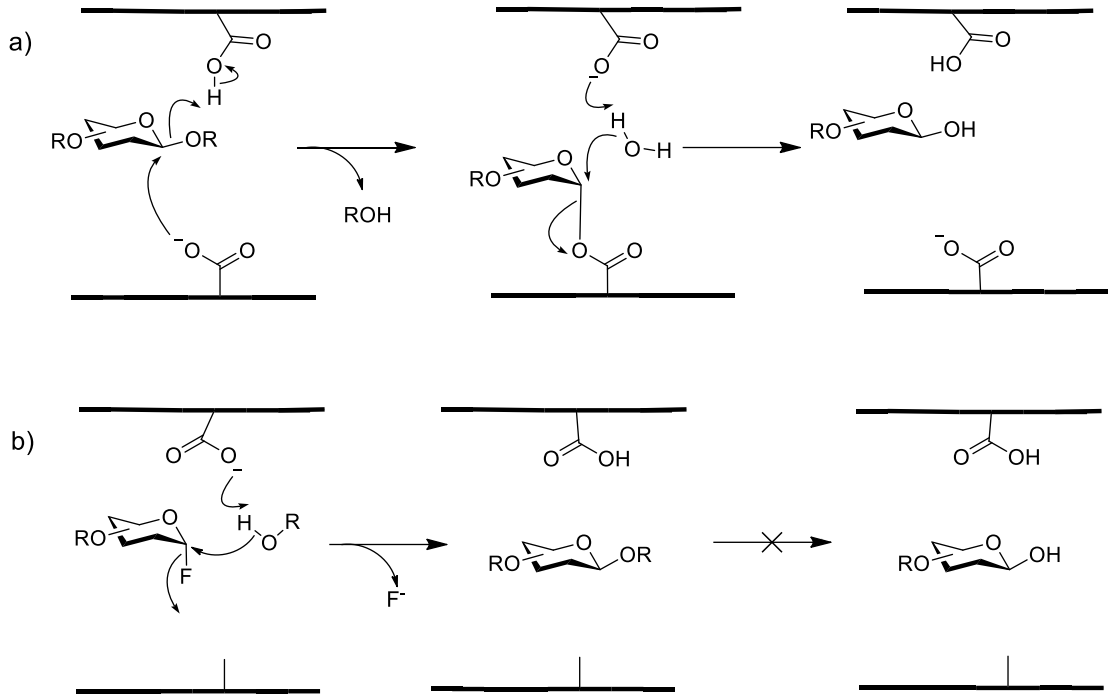


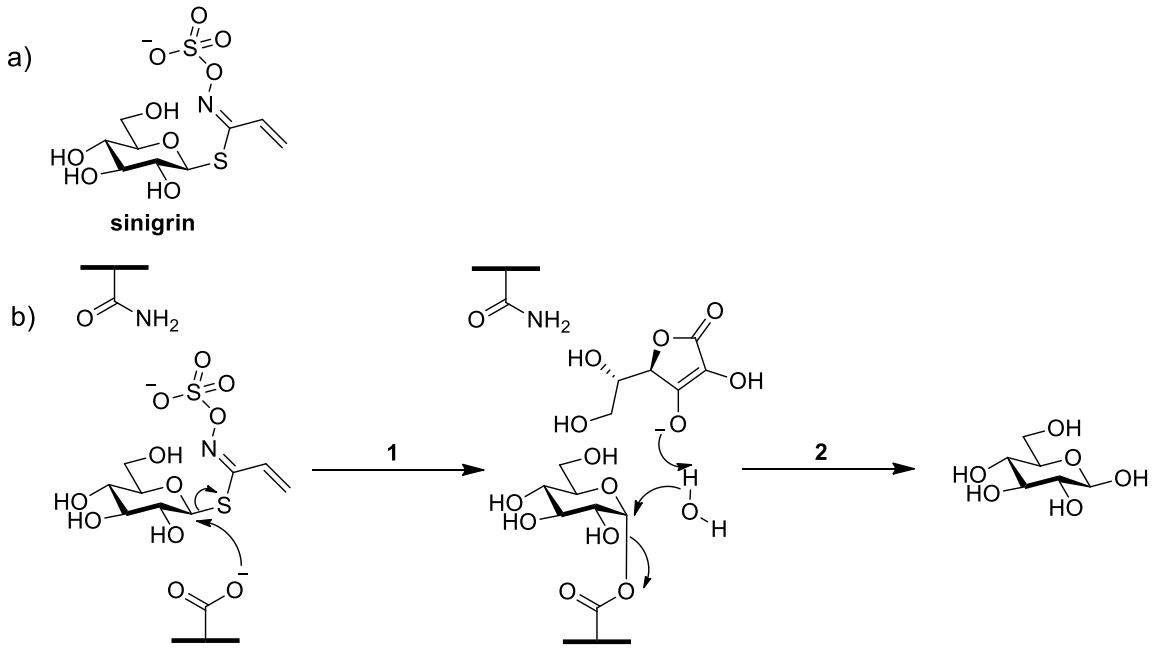
Figure 6 - Mechanisms of a) retaining β-glycosidase and; b) β-glycosynthase catalysis.

However, with an intact active site, it could still catalyse the ligation of an activated α-glycosyl donor to an acceptor bound in the aglycon pocket (Figure 6b). The enzyme, without nucleophilic activity, could not form the covalent glycosyl-enzyme intermediate, and so hydrolysis of the product did not take place. Withers termed these mutated enzymes glycosynthases, "mutant enzymes formed from retaining glycosidases that are capable of catalysing glycosyl transfer from glycosyl fluorides to acceptor alcohols without hydrolysis of the product".[95]

Since the invention of glycosynthases, there have been multiple developments from various research groups that have focused on screening a variety of different amino acid mutations (e.g. nucleophilic[96] and acid/base mutants[97]) as well as broadening substrate specificity to generate desired linkages.[97–99] However there is no 'general strategy' for their production.[100,101]

1.12 Chemical rescue of mutant glycosidases

The restoration of the activity of inactive mutant enzymes by small exogenous molecules is a well-established experimental strategy. In fact, the process has even been demonstrated in Nature by myrosinase, a glucosidase expressed in mustard and rapeseed.[102] Myrosinase lacks a glutamic acid residue at the position which would normally act as catalytic acid-base residue (Scheme 3). Instead, the high reactivity of the sinigrin substrate means the glycosylation step **1** can be performed without protonic assistance. The subsequent de-glycosylation step **2** is then performed by ascorbate (a common plant metabolite) acting as a catalytic base (pK_a 4.1).

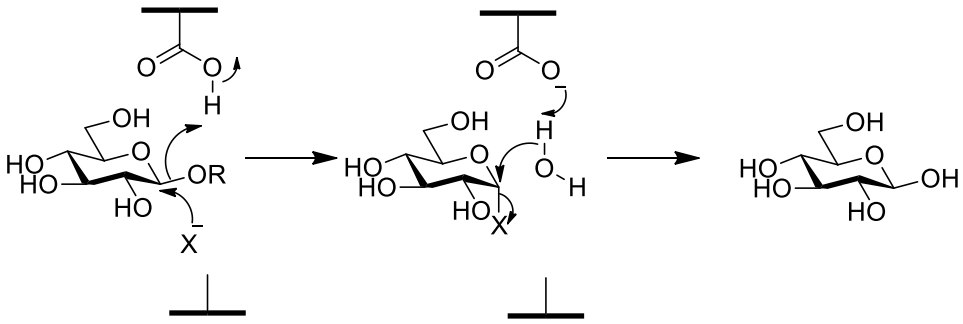


Scheme 3 – a) Chemical structure of sinigrin; b) Nature's version of chemical rescue found in myrosinase.

This methodology, so-called 'chemical rescue', has been used to probe the chemical and structural requirements for efficient catalysis by mutant enzymes.[76] In the cases of mutant glycosidases, chemical rescue has been used to identify important catalytic

residues in the enzyme active site.[103,104] As mentioned earlier, mutant glycosidases may have little hydrolytic activity on their respective natural substrates. However, in the presence of an activated substrate, such as a *p*-nitrophenyl or dinitrophenyl glycoside, and a small charged nucleophile such as azide, formate, or fluoride, the hydrolytic product can be formed.[75]

The mechanism of chemical rescue is illustrated on Scheme 4. Upon binding of the substrate by the enzyme, an exogenous nucleophile facilitates anomeric bond cleavage in place of the removed catalytic nucleophile. In such a case, the substitution of the catalytic nucleophile (commonly glutamate) of an enzyme for a smaller residue, such as glycine or alanine, provides a pocket in which the anion can bind. Chemical rescue is considered a diagnostic test for the identification of the nucleophilic residue.



Scheme 4 - Chemical rescue of mutant glycosidases using exogenous nucleophiles.

If the *in-situ* formed glycosyl fluoride is then reacted with an alcohol or a glycosyl acceptor, the "glycosynthase" reaction takes place.

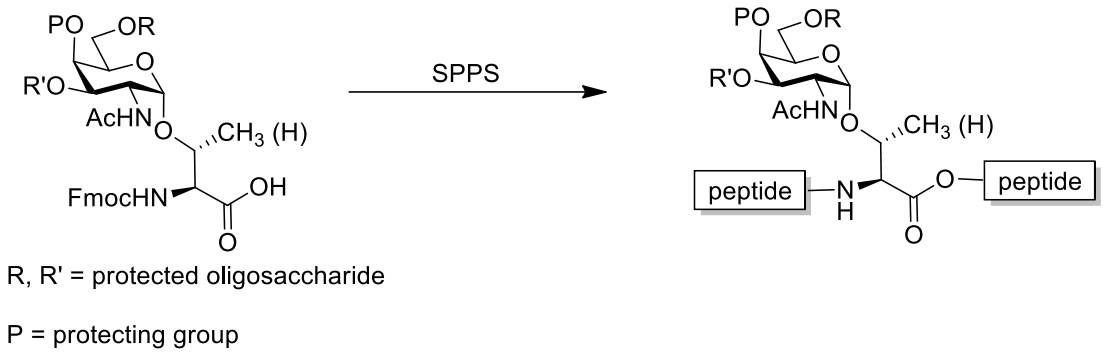
1.13 Methods of *O*-linked glycopeptide and glycoprotein synthesis

Glycoproteins are associated with a wide range of important biological processes. However, glycoproteins naturally exist as heterogenous mixtures, in which the same protein is linked to a variety of different oligosaccharides. This heterogeneity arises due to the fact that glycoprotein synthesis is not under direct genetic control. Each glycoform of the protein may have different physical properties. To access pure glycoforms of glycoproteins, the glycoscience community has resorted to chemical, biochemical, and chemo-enzymatic solutions.

As described earlier, mucin-type proteins are known to serve as important recognition elements that mediate a multitude of cell-cell interactions. The importance of *O*-glycosylation has inspired much interest in the synthesis of defined glycoproteins for structural and functional studies.[105] Since certain *O*-glycan structures are associated with tumour progression, there has also been a strong interest in the generation of glycopeptide-based tumour vaccines.[106]

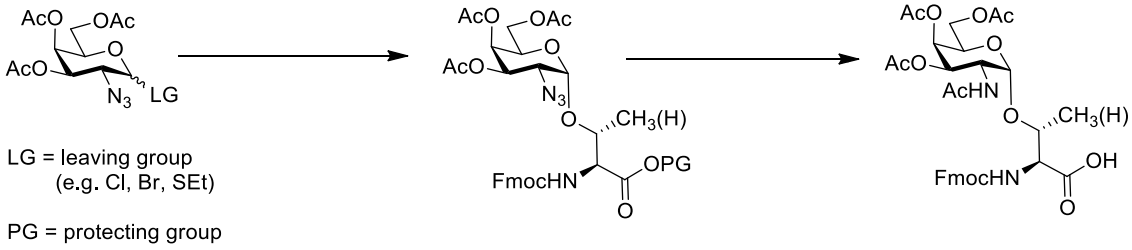
1.13.1. Solid-phase peptide synthesis (SPPS)

The most common chemical approach to the synthesis of glycopeptides with complex *O*-linked glycans involves the use of a suitably protected *O*-glycosyl amino acids as building blocks in solid-phase peptide synthesis (SPPS) (Scheme 5). In general, a base-labile amine protecting group, such as fluorenylmethyloxycarbonyl (Fmoc), is preferred over acid-labile protecting groups, such as *tert*-butoxycarbonyl (Boc)-based protection for the preparation of glycopeptides since basic conditions are more compatible with the presence of acid-sensitive anomeric glycosidic linkages.



Scheme 5 - General strategy for the synthesis of *O*-linked glycopeptides by SPPS.

Methods for the preparation of simple GalNAc-α-Ser/Thr building blocks for the assembly of mucin-type glycopeptides have been, for the most part, based on methodology introduced by Paulsen *et al.*[107] wherein α-GalNAc derivatives are prepared using a glycosyl donor that has a non-participating azido group at C-2 (Scheme 6). The resulting Fmoc-glycosyl amino acids can be incorporated into peptides using standard coupling procedures.[108]

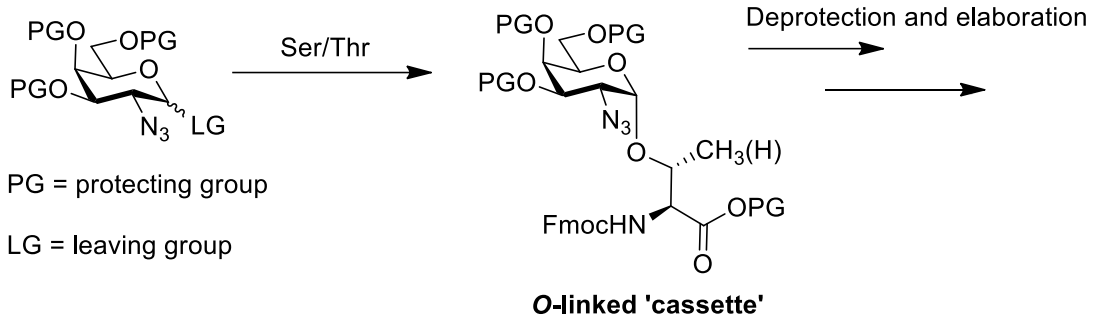


Scheme 6 - Synthesis of GalNAc building blocks using the methodology introduced by Paulsen *et al.*

1.13.2. Cassette Assembly

One of the main challenges in the synthesis of *O*-linked glycosyl amino acids is how to achieve high stereoselectivity in the formation of the α-glycosidic linkage between GalNAc and Ser/Thr, a problem that is even more pronounced when using oligosaccharide donors.[109] As a result, one of the most commonly employed methods

for the synthesis of complex *O*-glycosyl amino acids is using a ‘cassette assembly’ approach.[110] This strategy involves the initial formation of the desired α-Ser/Thr linkage prior to elaboration of additional sugars from the core GalNAc moiety (Scheme 7). Danishefsky and colleagues have used the cassette approach for the preparation of a variety of complex building blocks bearing tumour-related antigens.[111–113]

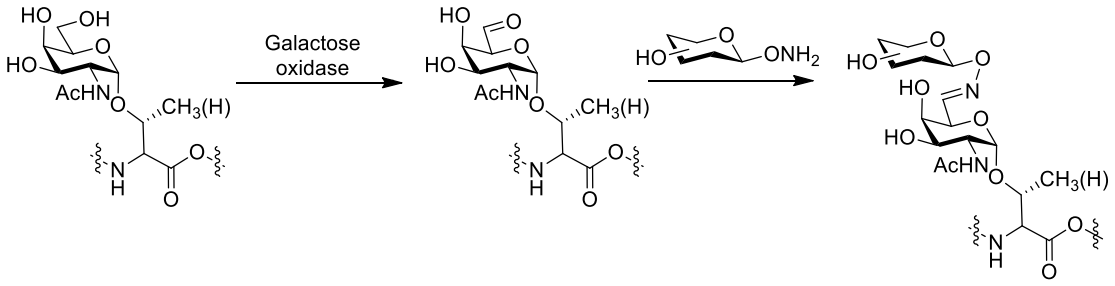


Scheme 7 - Generic schematic of the 'cassette approach'.

1.13.3. Chemoselective Ligation

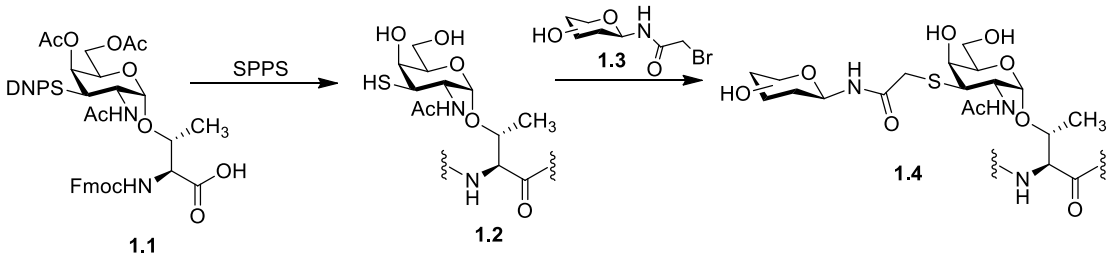
Chemoselective ligation is another technique that has been employed for the synthesis of *O*-linked glycopeptides. This method involves the attachment of preformed oligosaccharides to simple glycopeptides, and is an attractive method for the rapid assembly of peptides carrying complex *O*-glycans. Chemoselective ligation reactions are mild and selective, and allow for the coupling of unprotected biomolecules (such as peptides and carbohydrates) in an aqueous environment.[114–116] Rodriguez *et al.* used chemoselective ligation for the synthesis of *O*-linked glycopeptide mimetics by introducing unnatural bonds at the branch points (C-6) of the core GalNAc (Scheme 8).[117] A simple *O*-linked glycopeptide containing a single GalNAc unit was enzymatically oxidized using galactose oxidase to generate an aldehyde at C-6.

Chemoselective ligation with aminooxy-functionalized sugars gave the unnatural glycan containing an oxime linkage.



Scheme 8 - Synthesis of oxime-linked glycopeptide mimetics by chemoselective ligation.

Marcaurelle and Bertozzi have also used chemoselective ligation to generate glycopeptides with oligosaccharides attached to C-3 of the core GalNAc unit (Scheme 9).[118] Their synthesis started with building block **1.1** which contained a protected thiol at C-3. After glycosyl amino acid **1.2** was incorporated into a peptide by SPPS, the thiol was deprotected and then selectively alkylated with a variety of *N*-bromoacetamido glycosides **1.3** to generate thioether-linked glycopeptide mimetics **1.4**.

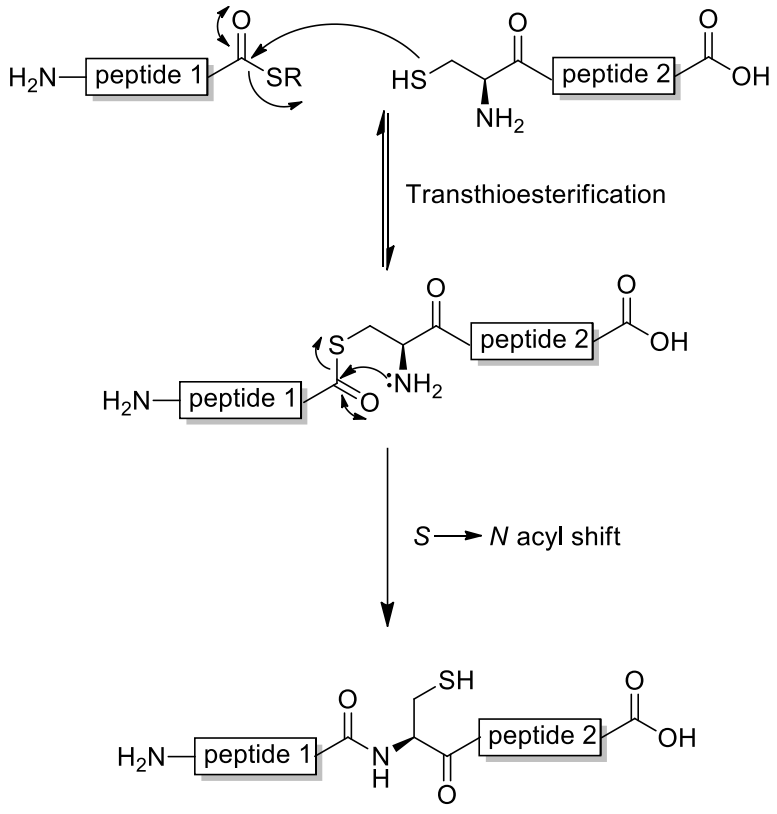


Scheme 9 - Synthesis of thioether-linked glycopeptide mimetics by chemoselective ligation.

1.13.4. Native chemical ligation

Native chemical ligation (NCL), was first applied to peptide synthesis in the 1990s by Kent and co-workers[119] and has since found widespread use in the field of protein chemistry for the synthesis of large un-glycosylated proteins.[120]

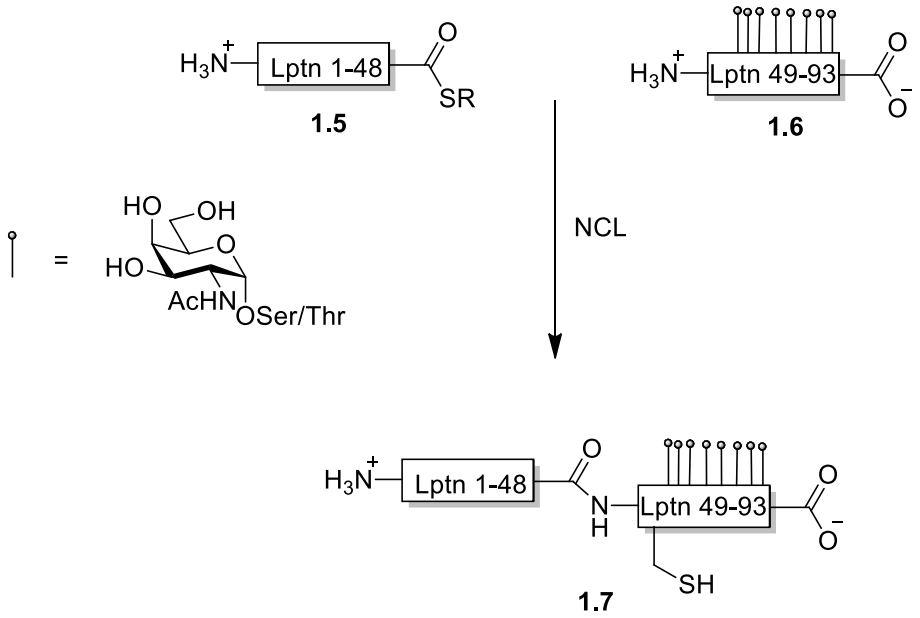
The ligation involves the reaction of an *N*-terminal cysteine of one peptide with the *C*-terminal thioester of another by transthioesterification (Scheme 10). Thioester formation is followed by a spontaneous S → N acyl shift to afford a native peptide bond. The NCL reaction is carried out in aqueous solution and is selective and often gives near-quantitative yields of the desired ligation products.



Scheme 10 - Mechanism of native chemical ligation.

In 2001, Marcaurelle *et al.* used NCL to synthesise the *O*-linked glycoprotein, lymphotactin (Lptn).[121] Lptn is a 93-amino-acid chemokine that acts as a potent

chemoattractant for T cells and natural killer cells.[122,123] The protein contains a small, mucin-like domain at the *C*-terminus which is *O*-glycosylated. The synthesis of Lptn was undertaken to investigate the structural and functional significance of the *O*-glycosylation. First, two fragments were synthesised (Scheme 11); a 47-residue peptide-α-thioester **1.5** and a 46-residue glycopeptide **1.6** containing 8 α-GalNAc residues. Ligation of fragments **1.5** and **1.6** by NCL was successful and gave the desired glycosylated chemokine **1.7** in milligram quantities.



Scheme 11 - Synthesis of glycosylated Lptn **1.7** by NCL.

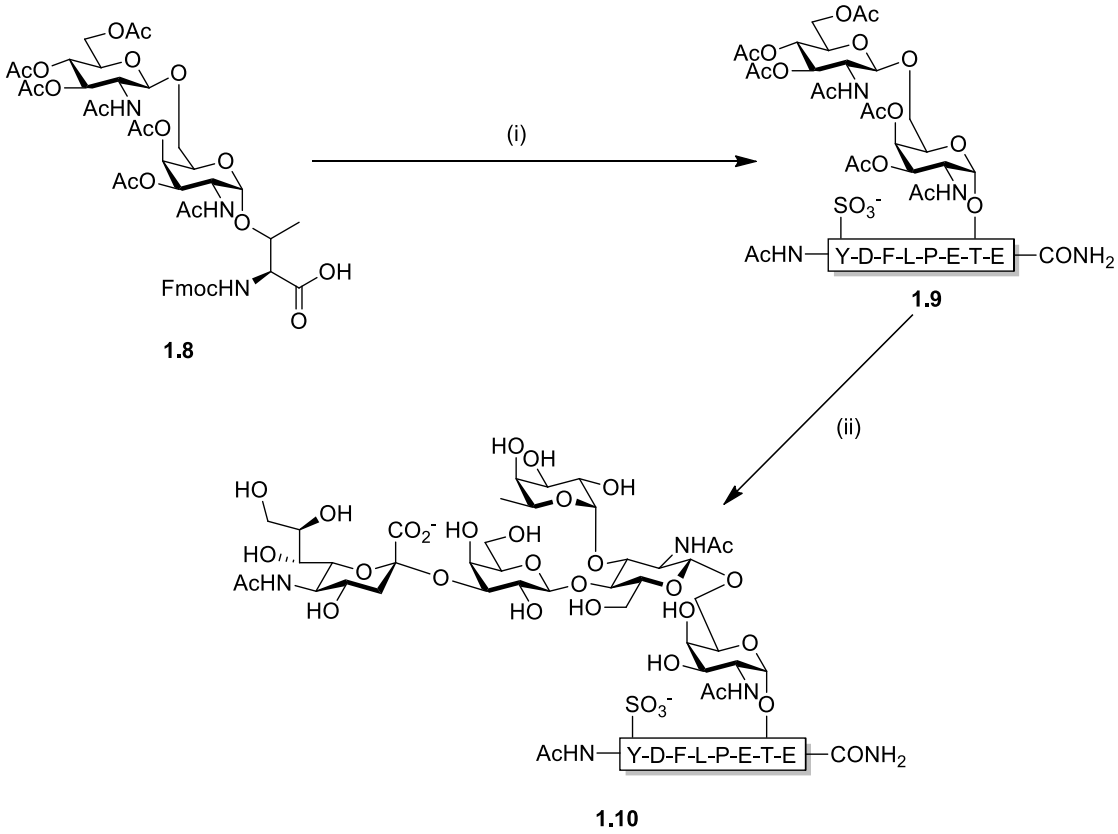
NCL has also been used to synthesise a number of important glycoproteins such as diptericin[124] and Ribonuclease C.[125,126]

1.13.5. Chemoenzymatic synthesis

Chemoenzymatic synthesis is a powerful and attractive alternative for the formation of glycosidic bonds under mild conditions without the need for protecting groups. The usage of glycosyltransferases for enzymatic transfer of individual monosaccharides to

preformed glycopeptides containing simple *O*-linked glycans is an attractive method for accessing large glycosyl amino acids.[127–131]

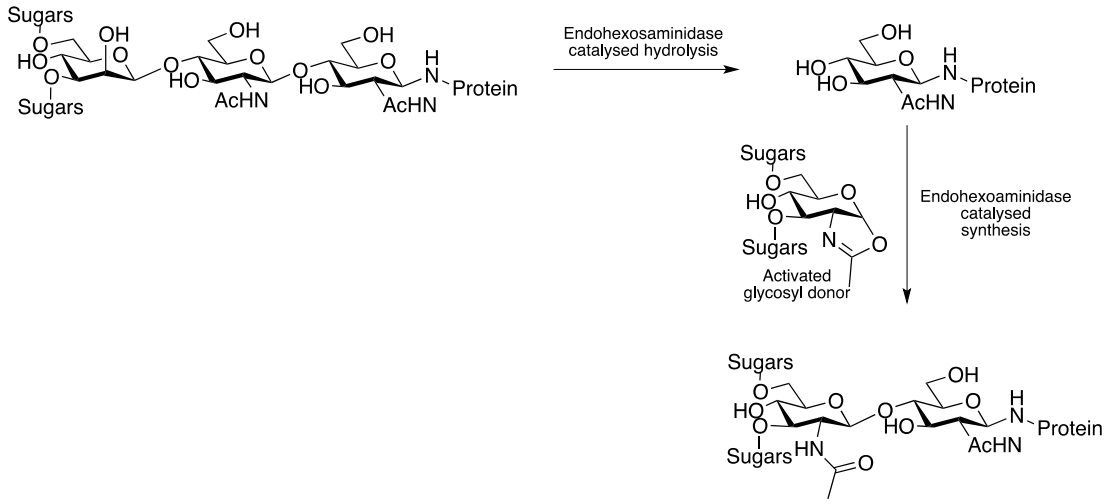
Koeller *et al*. have illustrated the elegance of this technique by the synthesis of a PSGL-1 fragment.[132] Herein, the synthesis of glycosyl amino acid **1.8** bearing an α-*O*-linked disaccharide was followed by the incorporation of **1.8** into a peptide using SPPS to give the synthetic glycopeptide **1.9** (Scheme 12). The disaccharide was then elaborated using a series of glycosyltransferases to give the desired PSGL-1 fragment **1.10**.



Scheme 12 - Synthesis of PSGL-1 **1.10** by Koeller *et al*.[133] Conditions: i) 1. SPPS, 2. SO_3·pyr, ii) 1. NaOMe, MeOH; 2. UDPGalβ1,4-GalT, 3) CMPNeuAcα2,3-SiaT, 4) GDPFucα1,3-FucT.

Glycosidases have also been successfully employed for glycopeptide synthesis. Particularly in the Fairbanks group, a technique known as glycoprotein remodelling has

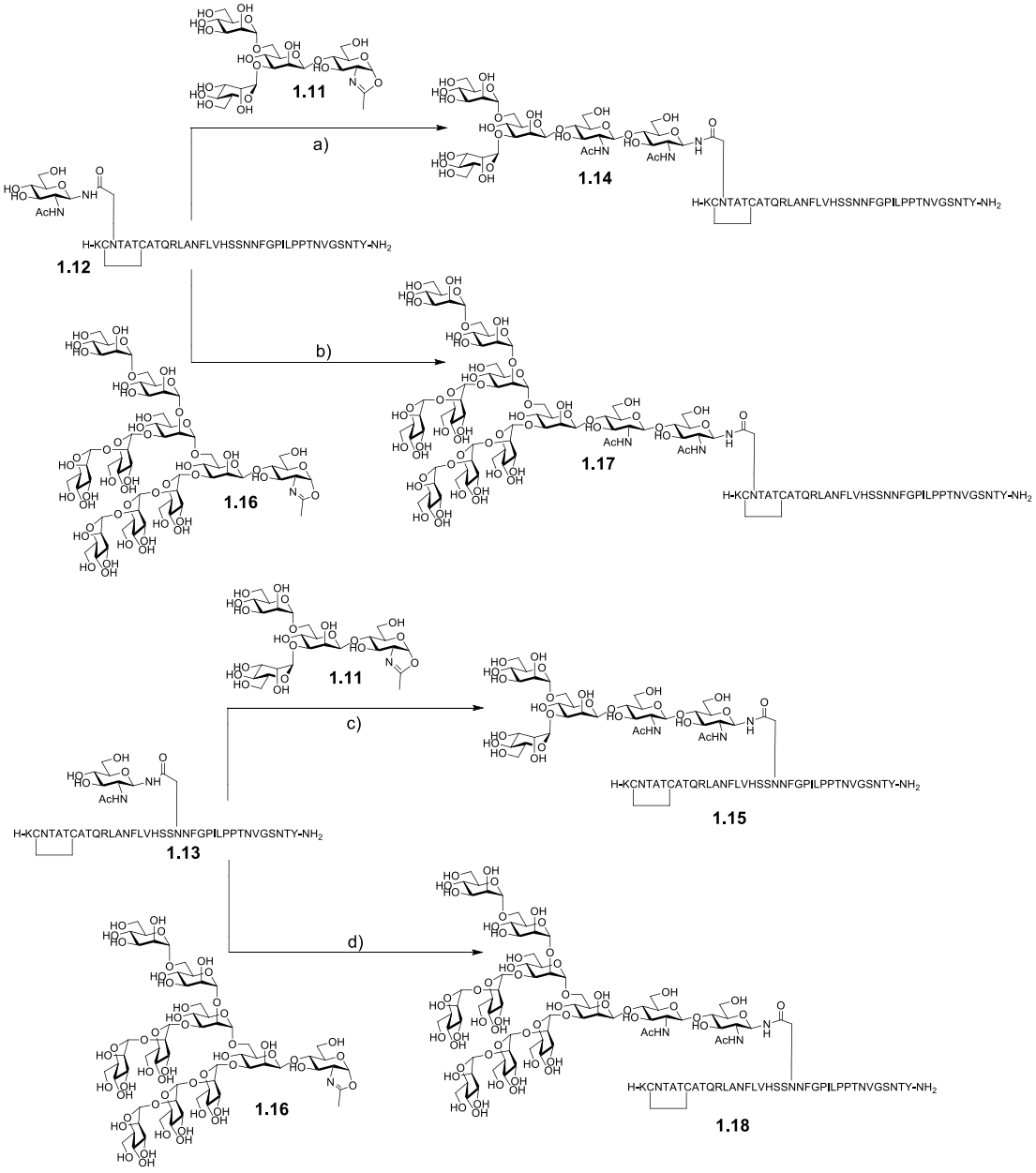
been used for the synthesis of glycoconjugates.[134–136] Glycoprotein remodelling, in the case of *N*-glycans, is a process wherein a heterogeneous mixture of glycoforms is trimmed down to a single sugar unit using endohexosaminidase enzymes, and is then enzymatically 'remodelled' by the attachment of defined oligosaccharides (Scheme 13).[136]



Scheme 13 - Endohexosaminidase-catalysed glycoprotein remodelling. The Fairbanks group have developed methods using Endo A from *Arthrobactor protomorphiae* and Endo M from *Mucor hiemalis* for glycoprotein remodelling.

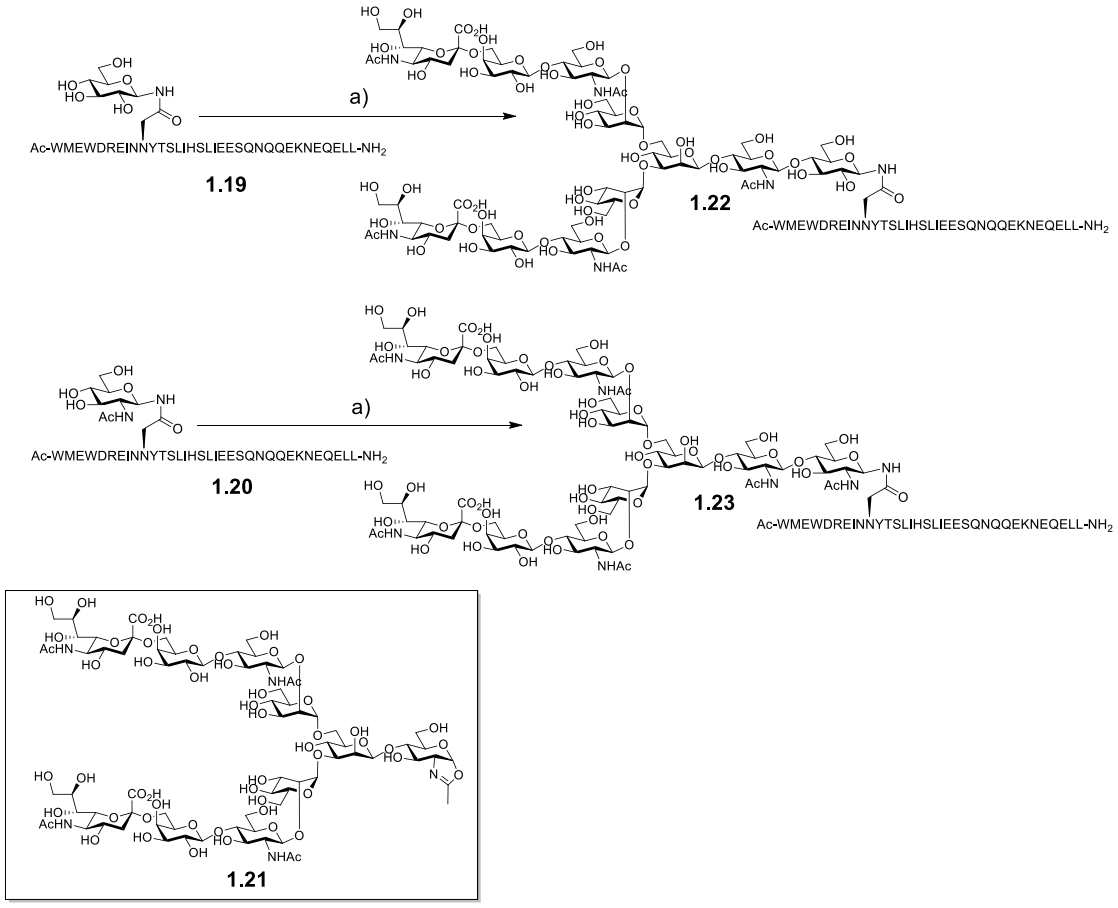
The use of glycosidases and glycosynthases for the synthesis of homogenous *N*-glycan glycopeptides and glycoproteins has extensively reported by the groups of Fairbanks[136,137] and Wang,[138,139] both groups whom have extensively used endohexosaminidases from the family GH18 and GH85. For example, Tomabechi *et al.*[140] synthesised glycosylated analogues of the diabetes drug pramlintide by a combination of solid-phase synthesis and enzymatic glycosylation (Scheme 14). The E173H[141] mutant of Endo A[142] was used to glycosylate tetrasaccharide oxazoline **1.11** to peptides **1.12** and **1.13** to give glycopramlintides **1.14** and **1.15** with the core *N*-glycan pentasaccharide attached. Moreover, the use of the N175Q mutant of Endo M

allowed the efficient attachment of the decasaccharide **1.16** to give undecasaccharide glycopramlintides **1.17** and **1.18**, with complex-type biantennary *N*-glycans at positions 3 and 21, respectively.



Scheme 14 - ENGase-catalyzed production of glycopramlintides with *N*-glycans at N3 and N23 by Tomabechi *et al.*[140] Conditions: a) Endo A E173H, sodium phosphate buffer, pH 6.5, 23 ˚C, 82 %; b) Endo M N175Q, sodium phosphate buffer, pH 6.5, 23 ˚C, 58 %; c) Endo A E173H, sodium phosphate buffer, pH 6.5, 23 ˚C, 95 %; d) Endo M N175Q, sodium phosphate buffer, pH 6.5, 23 ˚C, 96 %.

Lomino *et al.*[139] used Endo M N175A for the synthesis of a complex type glycoform of polypeptide C34, a potent HIV inhibitor derived from HIV-1 gp41 (Scheme 15).[143–145] The Glc- **1.19** and GlcNAc-peptide **1.20** acceptors were enzymatically glycosylated, using Endo M N175A, with the sialylated complex type *N*-glycan from SG-oxazoline **1.21** to give the corresponding glycosylation products **1.22** and **1.23**, respectively, with yields exceeding 80 % when additional enzyme was added.



Scheme 15 - ENGase-catalysed glycosylation of Glc-C34 and GlcNAc-C34 by Lomino *et al.*[139] Conditions: SG-oxazoline **1.21**, Endo M N175A, phosphate buffer, pH 7.4, 83 % (**1.22**), 85 % (**1.23**).

Although there are many endo- and exo-glycosidases that trim different *O*-glycan linkages, there is still to our knowledge no endoglycosidase that will cleave the β1,3 linkage of the reducing end sugar to provide a GalNAc glycosyl 'handle' for further remodelling. In fact, most research has been focused on enzymes which directly cleave

the α-Ser/Thr linkage.[88,146,147] As illustrated in Section 1.6, the multitude of diseases associated with erroneous *O*-glycosylation necessitates the synthesis of vaccines containing epitopes which stimulate an immune response. However, the methods used for their syntheses often result in multiple synthetic steps which inevitably leads to loss of material. Thus, there is an increasing need for syntheses that do not incorporate multi-step, protecting group methodologies whilst furnishing high yields and stereoselectivities.

Recently, several endoglycosidases that cleave extended *O*-glycan structures from the serine or threonine residues to which they are naturally attached were isolated. These enzymes, endo-α-*N*-acetylgalactosaminidases, belong to family GH101 of the glycoside hydrolases (http://www.cazy.org/), and demonstrated properties which showed promise for the enzymatic synthesis of homogeneous *O*-glycan glycopeptides and glycoproteins.

Therefore, our interest turned towards developing a methodology to use these enzymes for the chemoenzymatic synthesis of *O*-linked glycopeptides.

1.14 Endoglycosidases of the GH101 family.

The GH101 glycosyl hydrolase family currently contains proteins from bacterial species, most of which are from commensal human bacteria, although some may also be human pathogens. Since these enzymes naturally cleave *O*-glycan oligosaccharide structures from peptides and proteins, they can therefore, in theory, be used to catalyse the reverse reaction. To date, there are 148 enzymes in family GH101, in which 10 enzymes have been characterised, and three enzymes have had crystal structures solved.

In 2008, Koutsioulis *et al.*[148] identified putative endo-α-*N*-acetylgalactosaminidases using a BLAST search[149] using the protein sequence of engBL from *Bifidobacterium longum* JCM1217. Four potential genes were cloned and expressed: engEF from *Enterococcus faecalis*, engCP from *Clostridium perfringens*, engAA from *Arthrobacter aurescens,* and engPA from *Propionibacterium acnes*. The substrate specificity of the enzymes were then investigated. Table 1 summarises hydrolytic capabilities of the GH101 endo-α-GalNAcases.

	EngCP	EngEF	EngPA	EngSP	EngAA
	Product released (%)				
Galβ1,3GalNAcαpNP (core 1)	100	100	100	100	100
Galβ1,3(GlcNAcβ1,6)GalNAcαpNP (core 2)	2.5	2	0	0.6	0
GlcNAcβ1,3GalNAcαpNP (core 3)	6	100	100	3	27
Galβ1,3GlcNAcαpNP	0	0	0	0	0
GalNAcαpNP	4.4	2.2	1.8	1.2	30

Table 1 - Substrate specificity of GH101 endo-α-GalNAcases using pNP substrates.

These enzymes could all release p-nitrophenol from Galβ1,3GalNAc-α-pNP (core 1-pNP). EngEF and engPA were capable of hydrolysing GlcNAcβ1,3GalNAc-α-pNP (core 3-pNP), while engAA could only partially hydrolyse the core 3 disaccharide. EngAA was capable of hydrolysing GalNAc-α-pNP while the rest of the enzymes could only release traces of the monosaccharide. None of the enzymes could hydrolyse

Galβ1,3GlcNAc-α-pNP and low or no activity was detected when Galβ1,3(GlcNAcβ1,6)GalNAc-α-pNP (core 2 trisaccharide) was used as a substrate.

The enzymes could also release sugars from asialofetuin and mucin. The kinetic parameters of the family GH101 endo-α-GalNAcases using core 1-α-pNP and core 3-α-pNP were also calculated (Table 2).

		k_{cat} (s^{-1})	K_M (μM)
core 1-α-pNP	EngCP	19.9	70.93
	EngEF	51.17	47.85
	EngPA	2.009	3.781
	EngSP	10.51	40.37
	EngAA	25.89	33.87
	EngBF	17.8	21.8
core 3-α-pNP	EngEF	9.434	20.03
	EngPA	28.9	11.15

Table 2 - Kinetic parameters of endo-α-GalNAcases using core 1-α-pNP and core 3-α-pNP as substrates.

EngEF exhibited the highest k_{cat} when core 1-α-pNP was used as the substrate, while engPA had the lowest activity on that substrate with a k_{cat} approximate 25 times lower than that observed for engEF. Interestingly, engPA had a 3-fold higher k_{cat} for the core 3-α-pNP compared to engEF. None of the enzymes could hydrolyse Galβ1,3GlcNAc-α-pNP indicating that there was a strict preference for GalNAc as the glycon.

The transglycosylation activity of the family GH101 endo-α-GalNAcases was also investigated. EngCP, engEF, engPA, engSP and engAA were all incubated with core 1-α-pNP and various alkanols. Analysis of the reaction mixtures by t.l.c. indicated that all of the enzymes exhibited similar transglycosylation activity.

Two enzymes from family GH101 were chosen for further studies: engBL, from *Bifidobacterium longum*, had previously been characterised extensively in terms of sequence, structure, and kinetics. The enzyme demonstrated potential for further development.

EngEF, from *Enterococcus faecalis*, was capable of hydrolysing both core 1 and core 3 disaccharides from their substrates thus was chosen for the broader substrate specificity of the enzyme.

1.15 Endo-α-*N*-acetylgalactosaminidase from *Bifidobacterium longum* (engBL)

Bifidobacteria are thought to be key commensals that promote a healthy intestinal tract because of their many beneficial effects on the host, such as regulation of the state of the intestine, reduction of harmful bacteria and toxic compounds, immunomodulation, and anti-carcinogenic activity.[150–152] Bifidobacteria naturally colonise the lower intestinal tract, an environment that has a low abundance of mono- and disaccharides because such sugars are preferentially consumed by the host and microbes present in the upper intestinal tract. To survive, bifidobacteria produce various kinds of exo- and endoglycosidases in surface-bound and/or extracellular forms, which they can utilize to gain access to a diverse range of carbohydrates.[153]

In 2005, Fujita *et al*. discovered an enzyme that catalysed the liberation of galactosyl β1,3-*N*-acetyl-D-galactosamine (Galβ1,3GalNAc, core 1 disaccharide) α-linked to serine or threonine residues from mucin-type glycoproteins.[88] The enzyme was purified from the cultural filtrates of *Bifidobacterium longum* and called endo-α-*N*-acetyl-galactosaminidase (EC 3.2.1.97; endo-α-GalNAcase, glycopeptide α-*N*-acetylgalactosaminidase, engBF, engBL). EngBL is a protein consisting of 1,966 amino acid residues. The central domain (590-1381 amino acid residues) was found to exhibit 31-53% identity to proposed proteins of several bacteria including *Clostridium perfringens* and *Streptococcus pneumoniae*. The recombinant protein when expressed in *E. coli* was found to liberate Galβ1,3GalNAc (core 1) from Galβ1,3GalNAc-α-pNP (core 1-pNP) and asialofetuin, but did not release GalNAc, Galβ1,3(GlcNAcβ1,6)GalNAc (the core 2 trisaccharide), or GlcNAcβ1,3GalNAc (the core 3 disaccharide) from their respective *p*-nitrophenol substrates. An analysis of the amino acid residues of the enzyme that were critical for hydrolysis determined them to be Asp-682 and Asp-789.

It was also found that engBL was capable of glycosylating a variety of alcohol acceptors, and even sugars, with the core 1-pNP disaccharide (Table 3).

Ashida *et al*. furthered this work by glycosylating free serine and threonine, as well as various biologically active peptides with the core 1 disaccharide using engBL and found that the products were themselves completely hydrolysed.[154]

Acceptor	**Transfer Ratio[1]**
Methanol	53.7 %
Ethanol	46.8 %
1-Propanol	32.2 %
1-Butanol	34.9 %
1-Pentanol	28.3 %
1-Hexanol	8.1 %
1-Heptanol	2.0 %
1-Octanol	Trace
1-Nonanol	ND[2]
D-Glucose	19.7 %
D-Galactose	8.3 %
D-Mannose	2.5 %
L-Arabinose	18.0 %
Maltose	14.4 %

Table 3 - Acceptor specificity of engBL transglycosylation as reported by Fujita *et al.*[88]

Suzuki *et al.* reported the crystal structure of the native engBL protein at 2.0 Å resolution, and also performed docking and mutational analyses to elucidate the critical residues for substrate recognition by the GH101 endo-α-*N*-GalNAc-ases.[155] EngBL was found to be a monomeric protein (Figure 7). Superimposition of the structures of the GH101 enzymes, engBL, and engSP showed that both enzymes are remarkably similar. Of the seven domains assigned for the engSP structure, engBL contained domains 2-7, but lacked domain 1, which resembles a domain of the GH13 α-amylase family. A sequence alignment of domains 2-7 showed a 45.8 % sequence identity with several

insertions and deletions located at the periphery of the protein that do not appear to be related to substrate binding.

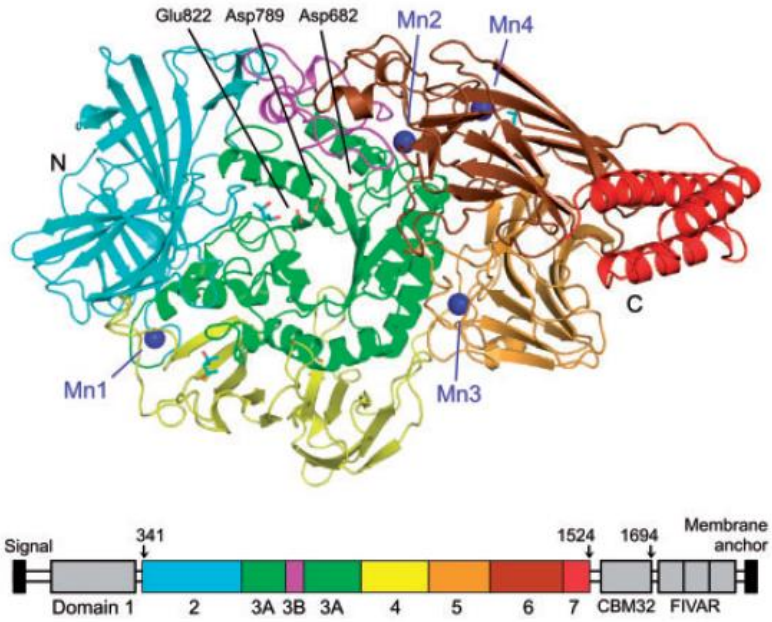


Figure 7 – The overall structure of engBL in a ribbon representation (top); and a schematic illustration of its domain organisation (bottom). Adapted from Suzuki *et al.*[155]

The residues in the active site of engSP that are involved in the catalysis were inferred to be Asp764 (nucleophile) and Glu796 (acid/base).[156] These residues correspond to Asp789 and Glu822 in engBL, which are located at the *C*-terminal loop of the fourth and fifth β-strands in domain 3A, respectively. Mutational studies at these positions demonstrated that catalytic activity was abolished.[88]

Autodocking analyses using Galβ1,3-GalNAc (GNB, galacto-*N*-biose) and Galβ1,3GlcNAc (LNB, lacto-*N*-biose) showed that the binding interactions for both substrates were the same. The anomeric C1 atom of GNB was found to be located in close proximity to the Oδ2 atom of Asp789 (3.1 Å), and the α-C1 hydroxyl group

pointed in the opposite direction from the nucleophile. The C1 hydroxyl group formed a hydrogen bond with the Oε2 atom of Glu822 (2.7 Å). Collectively, these results supported the mechanism of a retaining glycosidase.

Using mutagenesis, a third catalytically important residue was also identified, Asp682, which formed hydrogen bonds with three hydroxyl groups of the docked GNB (O4 and O6 of GalNAc and O6 of Gal), suggesting a substrate binding role for this residue.
The *N*-acetyl group of GalNAc fitted into a hydrophobic pocket formed by the side chains of His835, Trp836, and Tyr842, with the carbonyl oxygen forming a hydrogen bond with the side chain hydroxyl group of Tyr787. The side chain of Gln894 was located at subsite -2 and recognizes the C2 hydroxyl group of Gal.

Figure 8 shows the molecular surface of the substrate binding pocket of engBL and the partial sequence alignments of GH101 endo-α-GalNAc-ases.

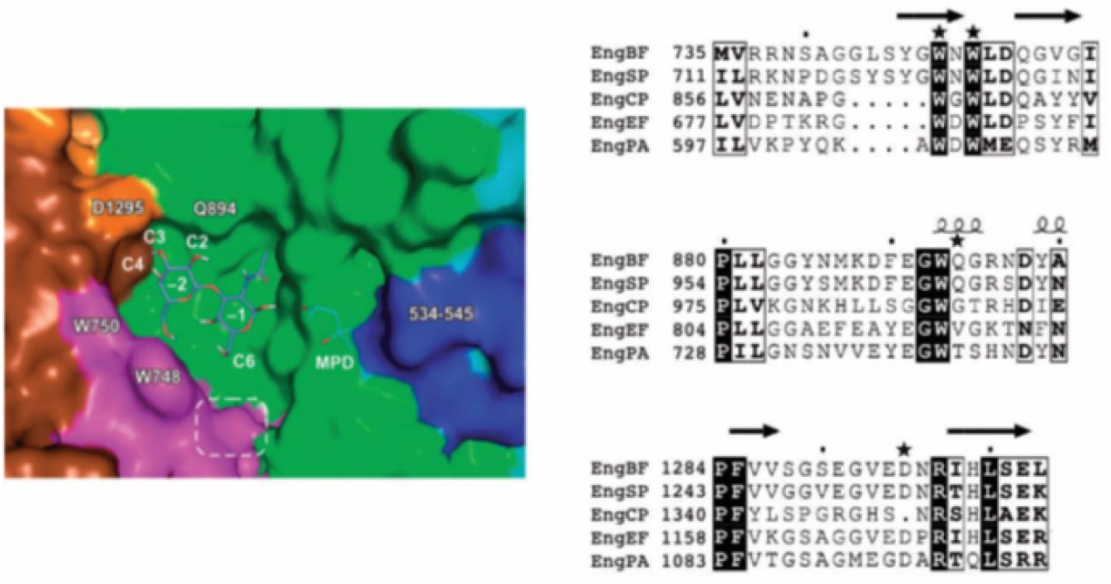
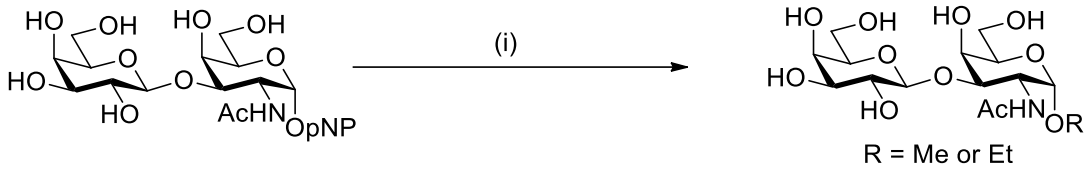


Figure 8 - Molecular surface of the substrate-binding pocket of engBL (left); and partial amino acid sequence alignment of GH101 endo-α-GalNAc-ases (right). Adapted from Suzuki *et al.*[155]

A protruded surface area formed by the *N*-terminal region adjacent to Trp748 (white dotted line on Figure 2, left) is observed near the C6 hydroxyl group of GalNAc at subsite -1. EngCP, engEF, and engPA have a deletion of four or five residues in this region allowing the accommodation of the branching β1-6 GlcNAc of the core 2 structure.[148]

1.16 Endo-α-*N*-acetylgalactosaminidase from *Enterococcus faecalis* (engEF)

In 2008, Goda *et al*. published an article detailing the activity of endo-α-*N*-acetylgalactosaminidase from *Enterococcus faecalis* (engEF), an ortholog of engBL.[92] The engEF gene was found to have 3972 base pairs corresponding to 1324 amino acids. The enzyme was capable of catalysing the hydrolysis of the core 1-pNP (Galβ1,3GalNAc-α-pNP), the core 2-pNP (Galβ1,3(GlcNAcβ1,6)GalNAc-α-pNP), the core 3-pNP (GlcNAcβ1,3GalNAc1-α-pNP), Gal-core 2-pNP, and GalNAc-α-pNP substrates. In the presence of methanol and ethanol, engEF was found to catalyse the transfer of the Galβ1,3GalNAc-α-pNP to generate the corresponding methyl and ethyl glycosides (Scheme 16).



Scheme 16 - Glycosylation of alkanols by engEF to core 1-pNP. Conditions: i) alkanol (15 % v/v), engEF from *Entercoccus faecalis*, 25 mM NaOAc buffer, pH 6.0, 37 °C.

Goda *et al*. summarised their findings regarding substrate specificity in five points:

1) The most preferred carbohydrate structure for the action of engEF was Galβ1,3GalNAc (core 1) amongst the substrates tested;
2) The terminal galactose of core 1 could be replaced by GlcNAc (core 3) but the rate of hydrolysis was decreased somewhat by this substitution;
3) The regiochemistry of the linkage of the terminal sugar in the core disaccharide structure was very important, i.e. the enzyme preferred a β1,3-linked sugar, not a β1,6-linked sugar;
4) Substitution of the OH-group at C6 of the internal GalNAc with GlcNAc (core 2) greatly reduced the activity, but this trisaccharide-pNP was hydrolysed by the enzyme and;
5) GalNAc-pNP and Gal-core 2-pNP were hydrolysed but at a slower rate than core 1-pNP or core 3-pNP.

To date, a crystal structure has not been published for engEF, and so the structural details that give rise to the broader specificity of this enzyme are not clear. However, as mentioned earlier, several residues normally clustered near the C6 hydroxyl of the substrate in engBL, were found to be absent in engEF which may allow for accommodation of the branching β1-6 GlcNAc unit.[148]

1.17 Project Objectives

The main goal of this research project was to develop a chemo-enzymatic method to synthesize *O*-glycan glycopeptides using family GH101 enzymes and their mutants.

This thesis describes the synthesis of activated *O*-glycan core structures and investigations into their use as donors for GH101 WT ENGase and mutant ENGase-

catalysed glycosylations. An additional chapter concerns investigations into the direct synthesis of glycosides in water.

Chapter 2 describes the transformation, expression, and purification of two enzymes from the family GH101, and their subsequent mutagenesis.

Chapter 3 details the total synthesis of activated *O*-glycan core structures (GalNAc, core 1, and core 3) as the *p*-nitrophenol and 1-*O*-acetyl derivatives that were to be used for the enzyme catalysed glycosylation reactions in Chapter 4.

Chapter 4 describes the investigations into the activity of the WT and mutant engBL and engEF enzymes from Chapter 2. Their kinetic parameters of hydrolysis and substrate specificity is quantified and their glycosylation ability is investigated.

Chapter 5 details investigates the use of various nucleophiles in the DMC-activated glycosylation of sugars in water.

Chapter 6 concludes this thesis with the experimental details for the work produced in Chapters 2-5.

1.18 References

(1) Dwek, R. A. *Chem. Rev.* **1996**, *96* (2), 683–720.

(2) Laine, R. A. *Glycobiology* **1994**, *4* (6), 759–767.

(3) Varki, A. *Glycobiology* **1993**, *3* (2), 97–130.

(4) Okajima, T.; Irvine, K. D. *Cell* **2002**, *111* (6), 893–904.

(5) Sharon, N.; Lis, H. *Essays Biochem.* **1995**, *30*, 59–75.

(6) Lasky, L. A. *Annu. Rev. Biochem.* **1995**, *64*, 113–139.

(7) Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A. *Science* **2001**, *291* (5512), 2370–2376.

(8) Helenius, A.; Aebi, M. *Science* **2001**, *291* (5512), 2364–2369.

(9) Wyss, D. F.; Choi, J. S.; Li, J.; Knoppers, M. H.; Willis, K. J.; Arulanandam, A R.; Smolyar, A.; Reinherz, E. L.; Wagner, G. *Science* **1995**, *269* (5228), 1273–1278.

(10) Aplin, J. D.; Wriston, J. C. *CRC Cr. Rev. Biochem. Mol.* **1981**, *10* (4), 259–306.

(11) Misaizu, T.; Matsuki, S.; Strickland, T. W.; Takeuchi, M.; Kobata, A.; Takasaki, S. *Blood* **1995**, *86* (11), 4097–4104.

(12) Rudd, P. M.; Joao, H. C.; Coghill, E.; Fiten, P.; Saunders, M. R.; Opdenakker, G.; Dwek, R. A. *Biochemistry* **1994**, *33* (1), 17–22.

(13) Yasomanee, J. P.; Demchenko, A. V. *J. Am. Chem. Soc.* **2012**, *134* (49), 20097–20102.

(14) Nigudkar, S. S.; Parameswar, A. R.; Pornsuriyasak, P.; Stine, K. J.; Demchenko, A. V. *Org. Biomol. Chem.* **2013**, *11* (24), 4068–4076.

(15) Yasomanee, J. P.; Demchenko, A. V. *Angew. Chem. Int. Ed.* **2014**, *53* (39), 10453–10456.

(16) Kaeothip, S.; Demchenko, A. V. *Carbohydr. Res.* **2011**, *346* (12), 1371–1388.

(17) Demchenko, A. V. *Synlett* **2003**, *9*, 1225–1240.

(18) Hasty, S. J.; Demchenko, A. V. *Chem. Heterocycl. Comp.* **2012**, *48* (1), 220–240.

(19) van den Steen, P.; Rudd, P. M.; Dwek, R. A.; Opdenakker, G. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33* (3), 151–208.

(20) Gowda, D. C.; Davidson, E. A. *J. Biol. Chem.* **1994**, *269* (31), 20031–20039.

(21) Kramerova, A. A.; Arbatsky, N. P.; Rozovsky, Y. M.; Mikhaleva, E. A.; Polesskaya, O. O.; Gvozdev, V. A.; Shibaev, V. N. *FEBS Lett.* **1996**, *378* (3), 213–218.

(22) Spiro, R. G. *Adv. Protein Chem.* **1973**, *27*, 349–467.

(23) Otvos, L.; Krivulka, G. R.; Urge, L.; Szendrei, G. I.; Nagy, L.; Xiang, Z. Q.; Ertl, H. C. J. *Biochim. Biophys. Acta - Mol. Cell. Res.* **1995**, 55–64.

(24) Pieper, J.; Ott, K. H.; Meyer, B. *Nat. Struct. Biol.* **1996**, 228–232.

(25) Kreppel, L. K.; Blomberg, M. A.; Hart, G. W. *J. Biol. Chem.* **1997**, *272* (14), 9308–9315.

(26) Lubas, W. A.; Frank, D. W.; Krause, M.; Hanover, J. A. *J. Biol. Chem.* **1997**,

272 (14), 9316–9324.

(27) Dong, D. L.; Hart, G. W. *J. Biol. Chem.* **1994**, *269* (30), 19321–19330.

(28) Gao, Y.; Wells, L.; Comer, F. I.; Parker, G. J.; Hart, G. W. *J. Biol. Chem.* **2001**, *276* (13), 9838–9845.

(29) Hanover, J. A.; Krause, M. W.; Love, D. C. *Nat. Rev. Mol. Cell Biol.* **2012**, *13* (5), 312–321.

(30) Zachara, N. E.; O'Donnell, N.; Cheung, W. D.; Mercer, J. J.; Marth, J. D.; Hart, G. W. *J. Biol. Chem.* **2004**, *279* (29), 30133–30142.

(31) Kim, E. Y.; Jeong, E. H.; Park, S.; Jeong, H.-J.; Edery, I.; Cho, J. W. *Genes Dev.* **2012**, *26* (5), 490–502.

(32) Zhu, Y.; Shan, X.; Yuzwa, S. A.; Vocadlo, D. J. *J. Biol. Chem.* **2014**, *289* (50), 34472–34481.

(33) Zhu, Y.; Liu, T.-W.; Cecioni, S.; Eskandari, R.; Zandberg, W. F.; Vocadlo, D. J. *Nat. Chem. Biol.* **2015**, *11* (5), 319–325.

(34) Hill, H. D.; Schwyzer, M.; Steinman, H. M.; Hill, R. L. *J. Biol. Chem.* **1977**, *252* (11), 3799–3804.

(35) Julenius, K.; Mølgaard, A.; Gupta, R.; Brunak, S. *Glycobiology* **2005**, *15* (2), 153–164.

(36) Yabu, M.; Korekane, H.; Miyamoto, Y. *Glycobiology* **2014**, *24* (6), 542–553.

(37) Cooper, C. A.; Wilkins, M. R.; Williams, K. L.; Packer, N. H. *Electrophoresis* **1999**, *20* (18), 3589–3598.

(38) Wang, C.; Fan, W.; Zhang, P.; Wang, Z.; Huang, L. *Proteomics* **2011**, *11* (21),

4229–4242.

(39) Jensen, P. H.; Karlsson, N. G.; Kolarich, D.; Packer, N. H. *Nat. Protoc.* **2012**, *7* (7), 1299–1310.

(40) Carlson, D. M. *J. Biol. Chem.* **1968**, *243* (3), 616–626.

(41) Hang, H. C.; Bertozzi, C. R. *Bioorg. Med. Chem.* **2005**, *13* (17), 5021–5034.

(42) Taylor-Papadimitriou, J.; Burchell, J.; Miles, D. W.; Dalziel, M. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1999**, *1455* (2-3), 301–313.

(43) Prokop, O.; Uhlenbruck, G. *Med. Welt* **1969**, *46*, 2515–2519.

(44) Springer, G. F.; Desai, P. R.; Banatwala, I. *Nature.* **1974**, *61* (10), 457–458.

(45) Numa, F.; Tsunaga, N.; Michioka, T.; Nawata, S.; Ogata, H.; Kato, H. *J. Obstet. Gynaecol.* **1995**, *21* (4), 385–389.

(46) Laack, E.; Nikbakht, H.; Peters, A.; Kugler, C.; Jasiewicz, Y.; Edler, L.; Hossfeld, D. K.; Schumacher, U. *Am. J. Pathol.* **2002**, *160* (3), 1001–1008.

(47) Konno, A.; Hoshino, Y.; Terashima, S.; Motoki, R.; Kawaguchi, T. *Clin. Exp. Metastasis* **2002**, *19* (1), 61–70.

(48) Fernández Madrid, F.; Tang, N.; Alansari, H.; Karvonen, R. L.; Tomkiel, J. E. *Autoimmun. Rev.* **2005**, *4* (4), 230–235.

(49) Kakeji, Y.; Tsujitani, S.; Mori, M.; Maehara, Y.; Sugimachi, K. *Cancer* **1991**, *68* (11), 2438–2442.

(50) Topaz, O.; Shurman, D. L.; Bergman, R.; Indelman, M.; Ratajczak, P.; Mizrachi, M.; Khamaysi, Z.; Behar, D.; Petronius, D.; Friedman, V.; Zelikovic, I.; Raimer, S.; Metzker, A.; Richard, G.; Sprecher, E. *Nat. Genet.* **2004**, *36* (6),

579–581.

(51) Wopereis, S.; Lefeber, D. J.; Morava, É.; Wevers, R. A. *Clin. Chem.* **2006**, *52* (4), 574–600.

(52) Berthier-Vergnes, O.; Zebda, N.; Bailly, M.; Bailly, C.; Doré, J. F.; Thomas, L.; Cochran, A. J. *Lancet*. **1993**, 1292.

(53) Zebda, N.; Bailly, M.; Brown, S.; Doré, J. F.; Berthier-Vergnes, O. *J. Cell. Biochem.* **1994**, *54* (2), 161–173.

(54) Hakomori, S. *Curr. Opin. Immunol.* **1991**, *3* (5), 646–653.

(55) Springer, G. F. *Science* **1984**, *224* (4654), 1198–1206.

(56) Ju, T.; Otto, V. I.; Cummings, R. D. *Angew. Chem. Int. Ed.* **2011**, *50* (8), 1770–1791.

(57) Nativi, C.; Renaudet, O. *ACS Med. Chem. Lett.* **2014**, *5* (11), 1176–1178.

(58) Sames, D.; Chen, X. T.; Danishefsky, S. J. *Nature* **1997**, *389* (6651), 587–591.

(59) Kim, Y. J.; Varki, A. *Glycoconj. J.* **1997**, *14* (5), 569–576.

(60) Blattman, J. N. *Science.* **2004**, *305* (5681), 200–205.

(61) Mitchell, M. S. *Int. J. Immunopharmacol.* **2003**, *3* (8), 1051–1059.

(62) Wilson, R. M.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2013**, *135* (39), 14462–14472.

(63) Yin, Z.; Chowdhury, S.; McKay, C.; Baniel, C.; Wright, W. S.; Bentley, P.; Kaczanowska, K.; Gildersleeve, J. C.; Finn, M. G.; BenMohamed, L.; Huang, X. *ACS Chem. Biol.* **2015**, *10* (10), 2364–2372.

(64) Richichi, B.; Thomas, B.; Fiore, M.; Bosco, R.; Qureshi, H.; Nativi, C.; Renaudet, O.; BenMohamed, L. *Angew. Chem. Int. Ed.* **2014**, *53* (44), 11917–11920.

(65) Springer, G. F.; Desai, P. R.; Tegtmeyer, H.; Spencer, B. D.; Scanlon, E. F. *Ann. N. Y. Acad. Sci.* **1993**, *690*, 355–357.

(66) Napoletano, C.; Rughetti, A.; Agervig Tarp, M. P.; Coleman, J.; Bennett, E. P.; Picco, G.; Sale, P.; Denda-Nagai, K.; Irimura, T.; Mandel, U.; Clausen, H.; Frati, L.; Taylor-Papadimitriou, J.; Burchell, J.; Nuti, M. *Cancer Res.* **2007**, *67* (17), 8358–8367.

(67) Kaiser, A.; Gaidzik, N.; Westerlind, U.; Kowalczyk, D.; Hobel, A.; Schmitt, E.; Kunz, H. *Angew. Chem. Int. Ed.* **2009**, *48* (41), 7551–7555.

(68) Westerlind, U.; Schröder, H.; Hobel, A.; Gaidzik, N.; Kaiser, A.; Niemeyer, C. M.; Schmitt, E.; Waldmann, H.; Kunz, H. *Angew. Chem. Int. Ed.* **2009**, *48* (44), 8263–8267.

(69) Hoffmann-Röder, A.; Kaiser, A.; Wagner, S.; Gaidzik, N.; Kowalczyk, D.; Westerlind, U.; Gerlitzki, B.; Schmitt, E.; Kunz, H. *Angew. Chem. Int. Ed.* **2010**, *49* (45), 8498–8503.

(70) Thompson, P.; Lakshminarayanan, V.; Supekar, N. T.; Bradley, J. M.; Cohen, P. A.; Wolfert, M. A.; Gendler, S. J.; Boons, G.-J. *Chem. Commun.* **2015**, *51* (50), 10214–10217.

(71) Abdel-Aal, A.-B. M.; Lakshminarayanan, V.; Thompson, P.; Supekar, N.; Bradley, J. M.; Wolfert, M. A.; Cohen, P. A.; Gendler, S. J.; Boons, G.-J. *ChemBioChem* **2014**, *15* (10), 1508–1513.

(72) Johannes, M.; Oberbillig, T.; Hoffmann-Röder, A. *Org. Biomol. Chem.* **2011**, *9* (15), 5541–5546.

(73) Unverzagt, C.; Kajihara, Y. *Chem. Soc. Rev.* **2013**, *42* (10), 4408.

(74) Niggemann, J.; Kamerling, J. P.; Vliegenthart, J. F. G. *Bioorg. Med. Chem.* **1998**, *6* (9), 1605–1612.

(75) Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G. *Annu. Rev. Biochem.* **2008**, *77*, 521–555.

(76) Zechel, D. L.; Withers, S. G. *Curr. Opin. Chem. Biol.* **2001**, *5* (6), 643–649.

(77) Koeller, K. M.; Wong, C.-H. *Chem. Rev.* **2000**, *100* (12), 4465–4494.

(78) Moretti, R.; Chang, A.; Peltier-Pain, P.; Bingman, C. A.; Phillips, G. N.; Thorson, J. S. *J. Biol. Chem.* **2011**, *286* (15), 13235–13243.

(79) Zhang, C.; Griffith, B. R.; Fu, Q.; Albermann, C.; Fu, X.; Lee, I.; Li, L.; Thorson, J. S. *Science.* **2006**, *313* (5791), 1291–1294.

(80) Williams, G. J.; Zhang, C.; Thorson, J. S. *Nat. Chem. Biol.* **2007**, *3* (10), 657–662.

(81) Gantt, R. W.; Peltier-Pain, P.; Cournoyer, W. J.; Thorson, J. S. *Nat. Chem. Biol.* **2011**, *7* (10), 685–691.

(82) CAZy Database (Carbohydrate-Active enZYmes) http://www.cazy.org/.

(83) Koshland, D. E. *Biol. Rev.* **1953**, *28* (4), 416–436.

(84) Withers, S. G.; Street, I. P. *J. Am. Chem. Soc.* **1988**, *110* (25), 8551–8553.

(85) McCarter, J. D.; Withers, S. G. *Curr. Opin. Struct. Biol.* **1994**, *4* (6), 885–892.

(86) Huang, X.; Tanaka, K. S. E.; Bennet, A. J. *J. Am. Chem. Soc.* **1997**, *119* (46), 11147–11154.

(87) Rye, C. S.; Withers, S. G. *Curr. Opin. Chem. Biol.* **2000**, *4* (5), 573–580.

(88) Fujita, K.; Oura, F.; Nagamine, N.; Katayama, T.; Hiratake, J.; Sakata, K.; Kumagai, H.; Yamamoto, K. *J. Biol. Chem.* **2005**, *280* (45), 37415–37422.

(89) Usui, T.; Matsui, H.; Isobe, K. *Carbohydr. Res.* **1990**, *203* (1), 65–77.

(90) Crout, D. H.; Vic, G. *Curr. Opin. Chem. Biol.* **1998**, *2* (1), 98–111.

(91) Mackenzie, L. F.; Wang, Q.; Warren, R. A. J.; Withers, S. G. *J. Am. Chem. Soc.* **1998**, *120* (22), 5583–5584.

(92) Goda, H. M.; Ushigusa, K.; Ito, H.; Okino, N.; Narimatsu, H.; Ito, M. *Biochem. Biophys. Res. Commun.* **2008**, *375* (4), 541–546.

(93) Moracci, M.; Trincone, A.; Perugino, G.; Ciaramella, M.; Rossi, M. *Biochemistry* **1998**, *37* (49), 17262–17270.

(94) Malet, C.; Planas, A. *FEBS Lett.* **1998**, *440* (1-2), 208–212.

(95) Williams, S. J.; Withers, S. G. *Carbohydr. Res.* **2000**, *327* (1-2), 27–46.

(96) Mayer, C.; Zechel, D. L.; Reid, S. P.; Warren, R. A. J.; Withers, S. G. *FEBS Lett.* **2000**, *466* (1), 40–44.

(97) Müllegger, J.; Jahn, M.; Chen, H. M.; Warren, R. A. J.; Withers, S. G. *Protein Eng. Des. Sel.* **2005**, *18* (1), 33–40.

(98) Jahn, M.; Marles, J.; Warren, R. A. J.; Withers, S. G. *Angew. Chem. Int. Ed.* **2003**, *42* (3), 352–354.

(99) Hancock, S. M.; Rich, J. R.; Caines, M. E. C.; Strynadka, N. C. J.; Withers, S. G. *Nat. Chem. Biol.* **2009**, *5* (7), 508–514.

(100) Honda, Y.; Kitaoka, M. *J. Biol. Chem.* **2006**, *281* (3), 1426–1431.

(101) Honda, Y.; Fushinobu, S.; Hidaka, M.; Wakagi, T.; Shoun, H.; Taniguchi, H.; Kitaoka, M. *Glycobiology* **2008**, *18* (4), 325–330.

(102) Burmeister, W. P.; Cottaz, S.; Rollin, P.; Vasella, A.; Henrissat, B. *J. Biol. Chem.* **2000**, *275* (50), 39385–39393.

(103) Cobucci-Ponzano, B.; Trincone, A.; Giordano, A.; Rossi, M.; Moracci, M. *Biochemistry* **2003**, *42* (32), 9525–9531.

(104) Ly, H. D.; Withers, S. G. *Annu. Rev. Biochem.* **1999**, *68*, 487–522.

(105) Herzner, H.; Reipen, T.; Schultz, M.; Kunz, H. *Chem. Rev.* **2000**, *100* (12), 4495–4538.

(106) Danishefsky, S.; Allen, J. *Angew. Chem. Int. Ed.* **2000**, *39* (5), 836–863.

(107) Paulsen, H.; Kolář, Č.; Stenzel, W. *Chem. Ber.* **1978**, *111* (6), 2358–2369.

(108) Han, S. Y.; Kim, Y. A. *Tetrahedron* **2004**, *60* (11), 2447–2467.

(109) Kalikanda, J.; Li, Z. *Carbohydr. Res.* **2011**, *346* (15), 2380–2383.

(110) Chen, X.-T.; Sames, D.; Danishefsky, S. J. *J. Am. Chem. Soc.* **1998**, *120* (31), 7760–7769.

(111) Kuduk, S. D.; Schwarz, J. B.; Chen, X.-T.; Glunz, P. W.; Sames, D.; Ragupathi, G.; Livingston, P. O.; Danishefsky, S. J. *J. Am. Chem. Soc.* **1998**, *120* (48), 12474–12485.

(112) Schwarz, J. B.; Kuduk, S. D.; Chen, X.-T.; Sames, D.; Glunz, P. W.; Danishefsky, S. J. *J. Am. Chem. Soc.* **1999**, *121* (12), 2662–2673.

(113) Glunz, P. W.; Hintermann, S.; Williams, L. J.; Schwarz, J. B.; Kuduk, S. D.; Kudryashov, V.; Lloyd, K. O.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2000**, *122* (30), 7273–7279.

(114) Lemieux, G.; Bertozzi, C. R. *Trends Biotechnol.* **1998**, *16* (12), 506–513.

(115) Marcaurelle, L. A.; Bertozzi, C. R. *Chem. Eur. J.* **1999**, *5* (5), 1384–1390.

(116) Hang, H. C.; Bertozzi, C. R. *Acc. Chem. Res.* **2001**, *34* (9), 727–736.

(117) Rodriguez, E. C.; Winans, K. A.; King, D. S.; Bertozzi, C. R. *J. Am. Chem. Soc.* **1997**, *119* (41), 9905–9906.

(118) Marcaurelle, L. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2001**, *123* (8), 1587–1595.

(119) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. *Science.* **1994**, *266* (5186), 776–779.

(120) Dawson, P. E.; Kent, S. B. H. *Annu. Rev. Biochem.* **2000**, *69* (1), 923–960.

(121) Marcaurelle, L. A.; Mizoue, L. S.; Wilken, J.; Oldham, L.; Kent, S. B. H.; Handel, T. M.; Bertozzi, C. R. *Chemistry.* **2001**, *7* (5), 1129–1132.

(122) Dorner, B.; Müller, S.; Entschladen, F.; Schröder, J. M.; Franke, P.; Kraft, R.; Friedl, P.; Clark-Lewis, I.; Kroczek, R. A. *J. Biol. Chem.* **1997**, *272* (13), 8817–8823.

(123) Hedrick, J. A.; Saylor, V.; Figueroa, D.; Mizoue, L.; Xu, Y.; Menon, S.; Abrams, J.; Handel, T.; Zlotnik, A. *J. Immunol.* **1997**, *158* (4), 1533–1540.

(124) Shin, Y.; Winans, K. A; Backes, B. J.; Kent, S. B. H.; Ellman, J. A; Bertozzi, C. R. *J. Am. Chem. Soc.* **1999**, *121* (50), 11684–11689.

(125) Piontek, C.; Ring, P.; Harjes, O.; Heinlein, C.; Mezzato, S.; Lombana, N.; Pöhner, C.; Püttner, M.; Varón Silva, D.; Martin, A.; Schmid, F. X.; Unverzagt, C. *Angew. Chem. Int. Ed.* **2009**, *48* (11), 1936–1940.

(126) Piontek, C.; Varón Silva, D.; Heinlein, C.; Pöhner, C.; Mezzato, S.; Ring, P.; Martin, A.; Schmid, F. X.; Unverzagt, C. *Angew. Chem. Int. Ed.* **2009**, *48* (11), 1941–1945.

(127) Takeuchi, H.; Kato, K.; Hassan, H.; Clausen, H.; Irimura, T. *Eur. J. Biochem.* **2002**, *269* (24), 6173–6183.

(128) Irimura, T.; Denda, K.; Iida, S. -i; Takeuchi, H.; Kato, K. *J. Biochem.* **1999**, *126* (6), 975–985.

(129) Kato, K.; Takeuchi, H.; Kanoh, A.; Mandel, U.; Hassan, H.; Clausen, H.; Irimura, T. *Glycobiology* **2001**, *11* (10), 821–829.

(130) Kagan, E.; Ragupathi, G.; Yi, S. S.; Reis, C. A.; Gildersleeve, J.; Kahne, D.; Clausen, H.; Danishefsky, S. J.; Livingston, P. O. *Cancer Immunol. Immunother.* **2005**, *54* (5), 424–430.

(131) Freire, T. *Glycobiology* **2006**, *16* (5), 390–401.

(132) Koeller, K. M.; Smith, M. E. B.; Huang, R.-F.; Wong, C.-H. *J. Am. Chem. Soc.* **2000**, *122* (17), 4241–4242.

(133) Koeller, K. M.; Smith, M. E. B.; Wong, C. -H. *Bioorg. Med. Chem.* **2000**, *8* (5), 1017–1025.

(134) Heidecke, C. D.; Parsons, T. B.; Fairbanks, A. J. *Carbohydr. Res.* **2009**, *344* (18), 2433–2438.

(135) Rising, T. W. D. F.; Heidecke, C. D.; Moir, J. W. B.; Ling, Z.; Fairbanks, A. J. *Chem. Eur. J.* **2008**, *14* (21), 6444–6464.

(136) Fairbanks, A. J. *Comptes Rendus Chim.* **2011**, *14* (1), 44–58.

(137) Fairbanks, A. J. *Pure Appl. Chem.* **2013**, *85* (9), 1847–1863.

(138) Fan, S. Q.; Huang, W.; Wang, L. X. *J. Biol. Chem.* **2012**, *287* (14), 11272–11281.

(139) Lomino, J. V.; Naegeli, A.; Orwenyo, J.; Amin, M. N.; Aebi, M.; Wang, L.-X. *Bioorg. Med. Chem.* **2013**, *21* (8), 2262–2270.

(140) Tomabechi, Y.; Krippner, G.; Rendle, P. M.; Squire, M. A.; Fairbanks, A. J. *Chem. Eur. J.* **2013**, *19* (45), 15084–15088.

(141) Heidecke, C. D.; Ling, Z.; Bruce, N. C.; Moir, J. W. B.; Parsons, T. B.; Fairbanks, A. J. *ChemBioChem* **2008**, *9* (13), 2045–2051.

(142) Takegawa, K.; Nakoshi, M.; Iwahara, S.; Yamamoto, K.; Tochikura, T. *Appl. Environ. Microbiol.* **1989**, *55* (12), 3107–3112.

(143) Chan, D. C.; Chutkowski, C. T.; Kim, P. S. *Proc. Natl. Acad. Sci.* **1998**, *95* (26), 15613–15617.

(144) Chan, D. C.; Kim, P. S. *Cell* **1998**, *93* (5), 681–684.

(145) Wang, L.-X.; Song, H.; Liu, S.; Lu, H.; Jiang, S.; Ni, J.; Li, H. *ChemBioChem* **2005**, *6* (6), 1068–1074.

(146) Willis, L. M.; Zhang, R.; Reid, A.; Withers, S. G.; Wakarchuk, W. W. *Biochemistry* **2009**, *48* (43), 10334–10341.

(147) Kiyohara, M.; Nakatomi, T.; Kurihara, S.; Fushinobu, S.; Suzuki, H.; Tanaka, T.; Shoda, S. I.; Kitaoka, M.; Katayama, T.; Yamamoto, K.; Ashida, H. *J. Biol. Chem.* **2012**, *287* (1), 693–700.

(148) Koutsioulis, D.; Landry, D.; Guthrie, E. P. *Glycobiology* **2008**, *18* (10), 799–805.

(149) Altschul, S. *Nucleic Acids Res.* **1997**, *25* (17), 3389–3402.

(150) Hooper, L. V. *Science.* **2001**, *292* (5519), 1115–1118.

(151) Gibson, G. R.; Wang, X. *J. Appl. Bacteriol.* **1994**, *77* (4), 412–420.

(152) Femia, A. P. *Carcinogenesis* **2002**, *23* (11), 1953–1960.

(153) Hoskins, L. C.; Agustines, M.; McKee, W. B.; Boulding, E. T.; Kriaris, M.; Niedermeyer, G. *J. Clin. Invest.* **1985**, *75* (3), 944–953.

(154) Ashida, H.; Ozawa, H.; Fujita, K.; Suzuki, S.; Yamamoto, K. *Glycoconj. J.* **2010**, *27* (1), 125–132.

(155) Suzuki, R.; Katayama, T.; Kitaoka, M.; Kumagai, H.; Wakagi, T.; Shoun, H.; Ashida, H.; Yamamoto, K.; Fushinobu, S. *J. Biochem.* **2009**, *146* (3), 389–398.

(156) Caines, M. E. C.; Zhu, H.; Vuckovic, M.; Willis, L. M.; Withers, S. G.; Wakarchuk, W. W.; Strynadka, N. C. J. *J. Biol. Chem.* **2008**, *283* (46), 31279–31283.

Chapter 2 – Expression, purification, and mutagenesis of family GH101 endo-α-*N*-acetylgalactosaminidases

2.1 Introduction

Enzymes have many advantages over conventional synthetic methods and their use is becoming more accepted in organic synthesis. Enzymes can be used without the requirement of a protecting group strategy whilst they maintain high stereo- and regioselectivity. In industrial processes, the use of enzymes is becoming increasingly widespread because of their catalytic efficiency, their ability to operate under 'green' conditions, and the ease of their production in large quantities through fermentation procedures.[1]

Over recent years, both glycosyl transferases and glycosyl hydrolases have been used extensively for the construction of glycosidic linkages.[2] As discussed in Chapter 1, the usage of glycosyl transferases for the construction of oligosaccharides is hampered by the lack of available enzymes, as well as the need for complex and expensive nucleotide donor substrates. On the other hand, there are glycosyl hydrolase enzymes (glycosidases), which naturally cleave the glycosidic linkage. However under controlled conditions, they can be used to catalyse the formation of glycosidic linkages.[3] The biocatalytic utility of glycosidases has been demonstrated on several occasions for the synthesis of various oligosaccharides.[4] The major issue associated with the usage of glycosyl hydrolases is that the product of the synthesis reaction is often also a substrate for the natural hydrolytic reaction[5,6], a point which led to the development of techniques that minimized product hydrolysis.[2,3,7] However, even with

these improvements, the yields of glycosidase-mediated glycosylations rarely exceeded 60 % and were not always regioselective.

A significant breakthrough occurred in 1998 when a new class of mutant glycosidases, so-called 'glycosynthases', was introduced.[8] The glycosynthase approach centred on the use of specifically mutated retaining glycosidases. In a normal hydrolysis reaction catalysed by a retaining glycosidase, the substrate initially binds to the enzyme, and this is then followed by the general acid-catalysed attack of nucleophilic residue in the enzyme active site at the anomeric centre to form a glycosyl-enzyme intermediate (Figure 9a). This covalent intermediate is then hydrolysed by general-base catalysed attack of water at the anomeric centre, forming a product with overall retention of configuration and returning the enzyme to its original protonation state.

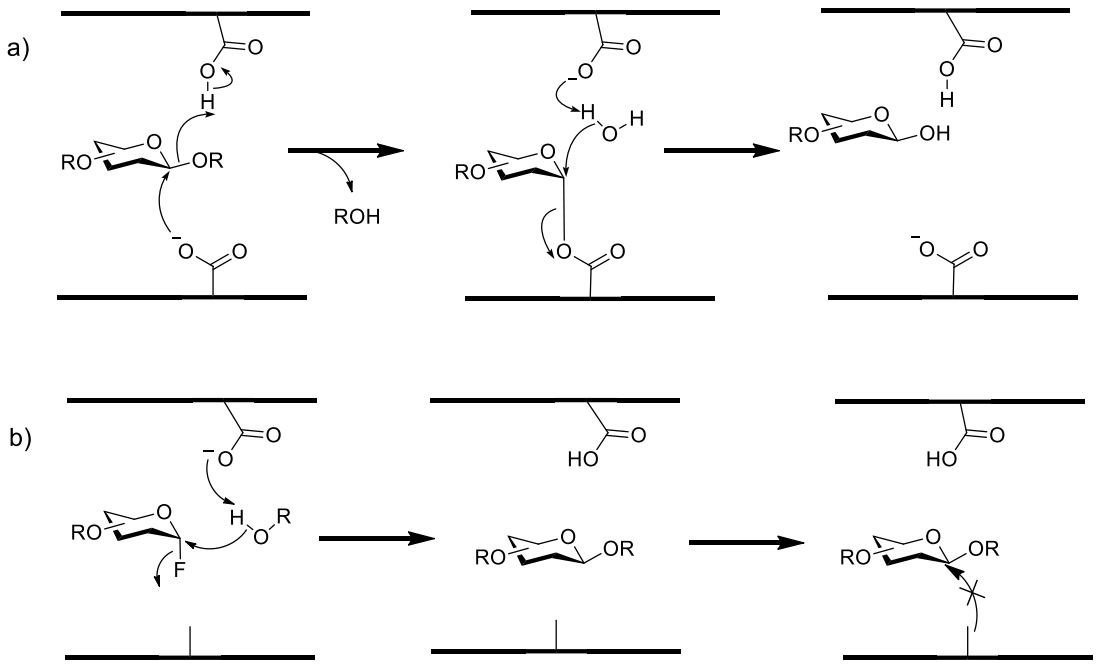


Figure 9 - Mechanism of a) β-retaining glycosidases and; b) glycosynthases.

In the original glycosynthase approach,[8] the catalytic nucleophile of a retaining glycosidase was substituted with a non-nucleophilic residue. The mutagenesis resulted in a correctly folded enzyme, but, without the ability to form the glycosyl-enzyme intermediate, the enzyme was catalytically inactive. However, in the presence of an activated glycosyl donor of the opposite anomeric configuration to the usual hydrolytic substrate for the enzyme (which mimics the α-glycosyl-enzyme intermediate), it was found that the enzyme could catalyse ligation of the donor to an acceptor substrate (Figure 9b). Moreover as the enzyme could not hydrolyse the product, this resulted in significantly increased synthetic efficiency. This initial discovery led to many subsequent efforts to engineer new glycosynthases to catalyse a variety of synthetic reactions .[1,9]

As described in Chapter 1, several endoglycosidases that cleave extended *O*-glycan structures from the serine or threonine residues to which they are naturally attached were recently isolated. These enzymes, endo-α-*N*-acetylgalactosaminidases, belong to super-family GH101 of the glycoside hydrolases,[10] and catalyse the hydrolysis of their respective substrates by means of a retaining mechanism as proposed by Koshland.[11] Retaining glycosidases use a double-displacement mechanism involving the formation of a glycosyl-enzyme intermediate. This intermediate is then hydrolysed by general base-catalysed attack of water upon the anomeric centre, forming the hemiacetal product with retention of configuration, and returning the enzyme to its original protonation state (Figure 10).

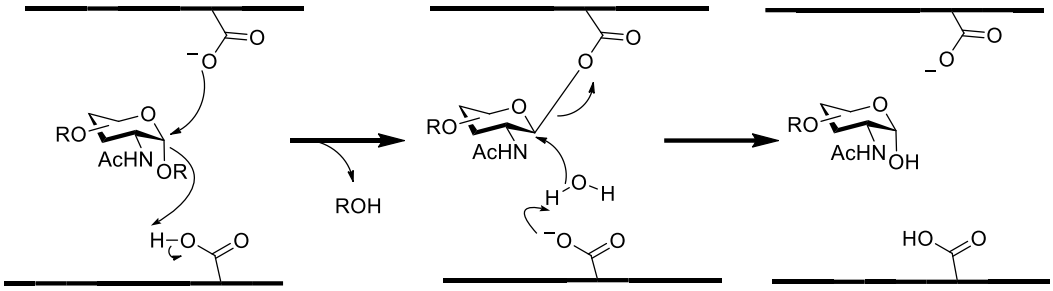


Figure 10 - Hydrolysis mechanism of family GH101 endo-α-*N*-acetylgalactosaminidases.

The enzymes demonstrated properties which showed promise for the enzymatic synthesis of homogeneous *O*-glycan glycopeptides and glycoproteins. Therefore, we were interested in glycoengineering glycosynthases from the GH101 hydrolases.

A standard approach for designing a glycosynthase begins with identifying the catalytic residues of interest in a target glycosidase. Online sequence alignment tools, such as ClustalW,[12] contain algorithms for comparing multiple sequences, and are useful for identifying residues of a protein involved in catalysis. Once a catalytic residue has been identified, the plasmid containing the gene coding for the enzyme can be mutated at the appropriate nucleotide position using site directed mutagenesis.

The mutant plasmid is then transformed, and the protein expressed, in a bacterial strain, commonly *E. coli*. The protein is then extracted, purified, and characterised by a variety of techniques, including circular dichroism (CD), and sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), to determine whether the protein has been correctly expressed and folded.

Two enzymes from the family GH101 were chosen for further studies: engBL, from *Bifidobacterium longum*, had previously been characterised extensively in terms of

sequence, structure, and kinetics.[13–15] The enzyme demonstrated potential for further development. EngEF, from *Enterococcus faecalis*, was capable of hydrolysing both Galβ1,3GalNAc (core 1) and GlcNAcβ1,3GalNAc (core 3) disaccharides from their substrates and thus was chosen for the broader substrate specificity of the enzyme.[16,17]

This chapter outlines the expression, purification, and mutagenesis of engBL and engEF of family GH101. Three mutations, namely glycine, alanine, and serine, were chosen, based on previous work on glycosynthase engineering[18,19], to replace the catalytic nucleophile of the respective enzyme.

The following sections serve to introduce the aforementioned molecular biology techniques to the reader before they are used in later sections of the chapter.

2.1.1 DNA Sequence Alignment using ClustalW

To date, the CAZY[10] database contains the sequences for well over 260,000 different glycosyl hydrolases, each which have been grouped into 135 families based on the DNA sequence analysis of the respective enzymes. DNA sequence analysis is an important bioinformatic technique for arranging multiple genetic data sets and identifying regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences. There are several online tools capable of performing the alignment of multiple sequences, however for the identification of catalytically important amino acid residues in engBL and engEF, ClustalW[12] was used.

ClustalW is a general purpose multiple alignment program for DNA and proteins. The tool operates by comparing multiple sequences and generating scores that quantify the similarity between pairs of sequences. These scores are then used to construct a dendrogram (guide phylogenetic tree) which is used as a basis for constructing the multiple sequence alignment whereby the most closely related pairs of sequences are aligned first.

For enzyme engineering, these tools are very useful for determining residues for mutagenesis. As demonstrated in Section 2.4.2., ClustalW was used for the identification of important catalytic amino acid residues and subsequent mutagenesis of the family GH101 enzyme, engEF.

2.1.2 Site Directed Mutagenesis

Mutagenesis of proteins is of paramount importance for understanding the relationships of protein structure and function. The functional and structural roles of amino acid residues in a protein of interest can be studied by comparing the mutant protein, carrying changes in amino acid residues, to the WT protein. However, to obtain specific mutant proteins, the mutant gene must be isolated or created.

Site-directed mutagenesis (SDM) is an invaluable technique for characterizing the dynamic, complex relationships between protein structure and function, for studying gene expression elements, and for carrying out vector modification. Three essential components are required in an *in vitro* DNA synthesis reaction: the template DNA, the primer, and the four deoxynucleotide triphosphates (dNTPs, Figure 11) in the presence of the enzyme, DNA polymerase.

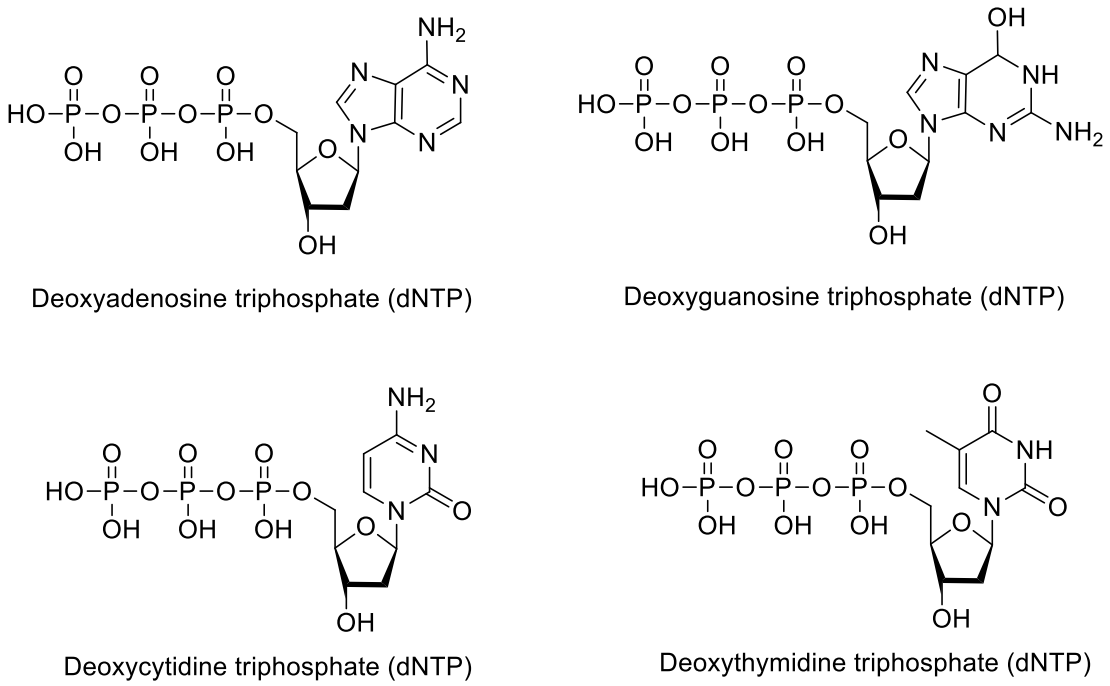


Figure 11 - Nucleotide triphosphates used in site-directed mutagenesis.

Primers are short strands of nucleic acid sequences that serve as a starting point for DNA synthesis, which is required since DNA polymerases can only add new nucleotides to an existing strand of DNA. Oligonucleotide primers are custom designed to incorporate a desired mutation in a double-stranded DNA plasmid. To ensure successful DNA mutagenesis, general rules for designing a good primer should be followed:

- A primer for DNA synthesis should be free of strong secondary structures, such as hairpins and direct repeats.

- The desired mutation, in general, should be placed in the middle of the primer rather than at the 5’ terminus. Mismatches at the 3’-terminus of a primer are known to affect rates of DNA synthesis.[20] The QuikChange Site-Directed

Mutagenesis Kit (Stratagene) protocol recommends that the desired mutation be placed in the middle of the primer sequence with 10-15 nucleotides flanking the mutation.

- To hybridize the mutagenic primer correctly and sufficiently to its template, the annealing temperature is chosen to be a few degrees below the melting temperature (T_m, the temperature at which half of the oligonucleotides are annealed to the template) of an oligonucleotide. In general, primers should have between 25-45 nucleotides in length with a T_m of 78 ˚C. The T_m can be estimated using the equation below, where N is the length of the primer:

$$T_m = 81.5 + 0.41*(\%GC) - \frac{675}{N} - \%\ \text{mismatch}$$

- Primers should also have a GC content of at least 40 %.

The design of primers generally follows a general procedure, whereby the DNA sequence of interest is identified, for example:

H D L D R L S Y L

5’ C A T G A C C T A G A T C G A C T T A G C T A T C T A 3’

3’ G T A C T C G A T C T A G C T G A A T C T A T A G A T 5’

The codon of interest (highlighted in red) is substituted for the one which encodes the desired mutation using a codon table (Figure 12).

First base	Second base								
		T		C		A		G	
	T	TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys
		TTC		TCC		TAC		TGC	
		TTA	Leu	TCA		TAA	STOP	TGA	STOP
		TTG		TCG		TAG		TGG	Trp
	C	CTT	Leu	CTT	Pro	CAT	His	CGT	Arg
		CTC		CCC		CAC		CGC	
		CTA		CCA		CAA	Gln	CGA	
		CTG		CCG		CAG		CGG	
	A	ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser
		ATC		ACC		AAC		AGC	
		ATA		ACA		AAA	Lys	AGA	Arg
		ATG	Met	ACG		AAG		AGG	
	G	GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly
		GTC		GCC		GAC		GGC	
		GTA		GCA		GAA	Glu	GGA	
		GTG		GCG		GAG		GGG	

Figure 12 - DNA Codon Table adapted from Garrett and Grisham, *Biochemistry*, 2nd ed.; Saunders College Publishing: F. L., 1999.

For example, a substitution of lysine (L) for valine (V):

H D L D R H S Y L

5’ C A T G A C C T A G A T C G A G T T A G C T A T C T A 3’

3’ G T A C T C G A T C T A G C T C A A T C T A T A G A T 5’

Short strand primers complimentary to the sequence above are then designed:

5’ A T A G C T A A C T C G A T C 3’ (complimentary to top strand)

5’ G A T C G A G T T A G C T A T 3’ (complimentary to bottom strand)

The Quikchange protocol employs the complementary primer pairs in the above steps in the polymerase chain reaction (PCR). The use of complementary primer pairs may lead to the formation of 'primer dimers', whereby partial annealing of a primer with a second primer in a reaction, due to stretches of complimentary bases, leads to the formation of tandem repeats of primers, reducing the yield of a successful transformation (Figure 13). To avoid this problem, various alternative protocols have been developed.[21–23]

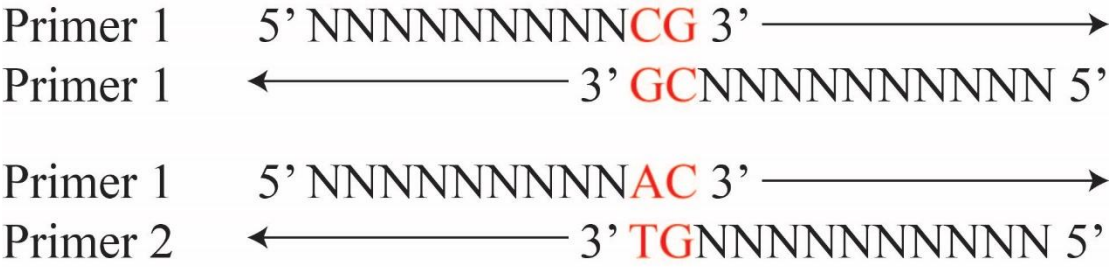


Figure 13 - The formation of primer dimers in PCR.

One drawback of using PCR for DNA mutagenesis is the relatively high rate of sequence errors in PCR products, often creating undesired mutations in addition to intended ones. *Taq* polymerase, from *Thermus aquaticus*, is the most widely used thermostable polymerase, however the enzyme lacks the 3'-5' exonuclease activity that "proof-reads" any errors caused by the 5'-3' DNA polymerase during DNA synthesis. The absence of proof-reading capabilities in a polymerase can lead to the accumulation of errors through many cycles of amplification. Alternatively, high-fidelity, non-strand-displacing DNA polymerases, such as *Pfu* polymerase (from *Pyrococcus furiosus*), which have superior thermostability and proof-reading properties, are used to amplify the entire plasmid. *Pfu* polymerase has been shown to provide up to 15 times the sequence fidelity of *Taq*.[24]

The product of the reaction is never used as a template, meaning that the amplification is linear, as opposed to exponential as observed in standard PCR.

A cycle of the mutagenesis reaction begins with the heating the template DNA until the strands are separated (Figure 14). Upon cooling of the reaction mixture, the mutagenic primers anneal to their complimentary sequences on the template sequence. The DNA polymerase then binds to and extends the primers to generate a full copy of the template sequence including the desired mutation. Several repeats of the thermocycling reaction yields several copies of the mutant gene.

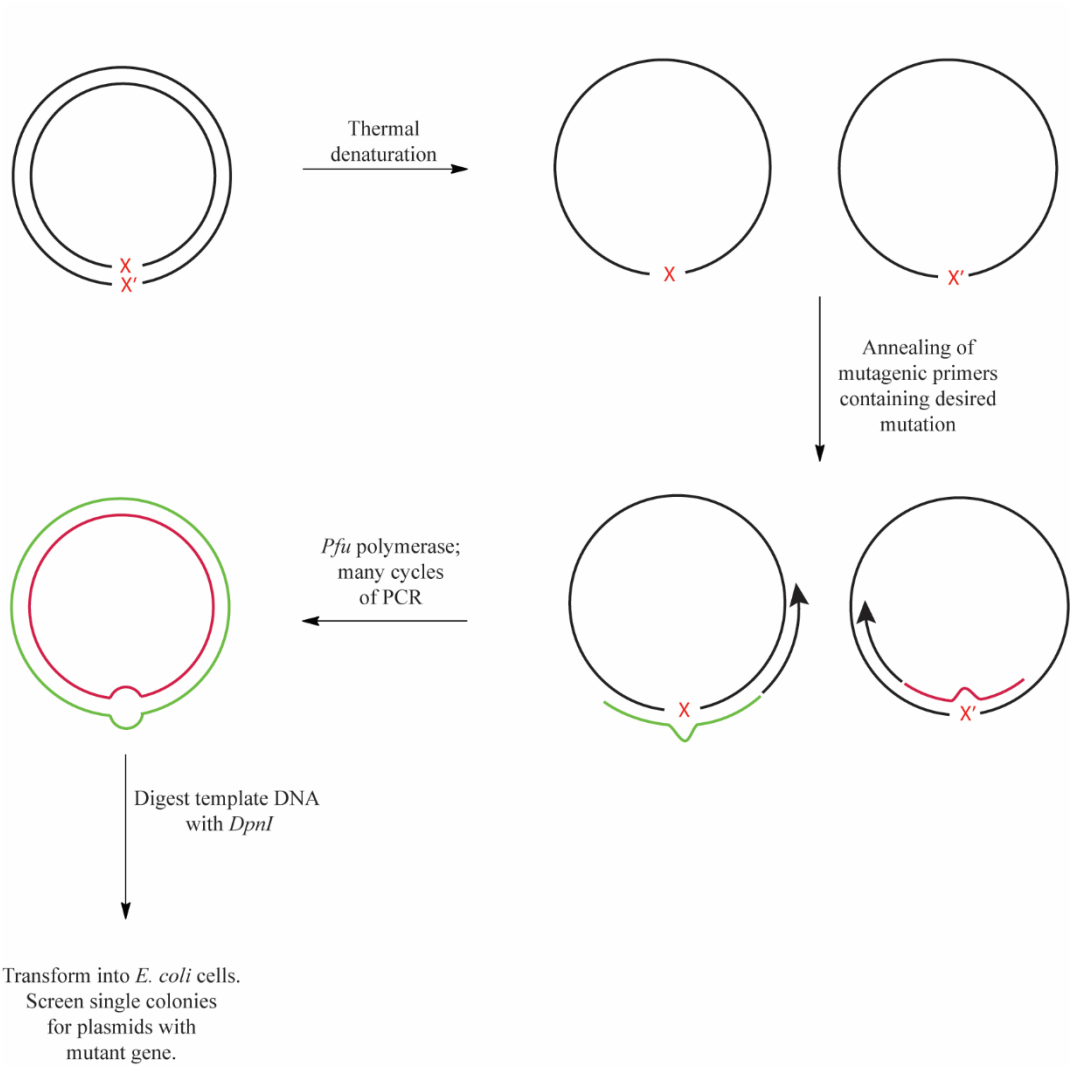


Figure 14 - Steps of site-directed mutagenesis, where X and X' represents the position of the desired mutation.

To prevent its transformation into *E. coli* and increase the likelihood of isolating the mutant gene, the template DNA is then enzymatically digested by *DpnI*, a restriction endonuclease, which is specific for methylated DNA (target sequence: 5'Gm^{6}ATC-3').[25] The mutated plasmid generated *in vitro* is unmethylated and is therefore undigested. The mutant nicked plasmid is then transformed into a bacterial strain, commonly *E. coli*, where it is ligated by ligase enzymes to form a circular plasmid. After the growth of the bacteria overnight, the plasmids are isolated using a plasmid preparation kit and sequenced to confirm the insertion of the desired point substitution.

2.1.3 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE is a common technique used to check the purity of protein samples. In the presence of the ionic detergent sodium dodecyl sulphate (SDS), proteins are denatured and the entire length of the protein is coated with an even distribution of negative charge along the protein chain due to the anionic properties of SDS (Figure 15).

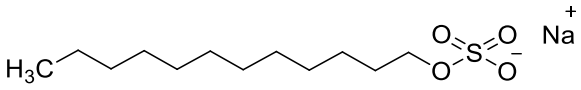


Figure 15 - Structure of sodium dodecyl sulfate

The purpose of this charge is multiple-fold. Firstly, since like charges repel, the proteins remain linear. Secondly, the even distribution of charges along the polypeptide chain results in a fractionation by approximate size during electrophoresis since the mobility of the protein depends only on the length and mass-to-charge ratio of the protein. Some quaternary structure may remain due to disulfide bonds and therefore a reducing agent (such as dithiothreitol (DTT) or β-mercaptoethanol) is added which breaks all disulfide bonds in the protein. Samples are normally heated to 70 – 90 ˚C for 10 mins before loading onto the gel, which completes the denaturing process. Weight size markers containing proteins of known weights are normally added to the gel in a separate lane to calibrate the gel and indicate the approximate molecular mass of unknown biomolecules by comparing the distance travelled relative to the marker.

An electric field is applied to the gel, causing the negatively charged proteins to migrate across the cell away from the negative electrode towards the positive electrode. Depending on their size, each biomolecule moves at a different rate through the gel matrix. Small molecules navigate more rapidly through the pores of the gel, whilst

larger molecules move more slowly. After allowing the proteins to run through the gel for a set amount of time, the gel is then stained, most commonly with Coomassie Brilliant Blue, allowing visualisation of the separated proteins.

2.1.4 Immobilised metal affinity chromatography

Immobilised metal affinity chromatography (IMAC) is a separation and purification technique based on the affinity of amino acids, particularly histidine, to metals. Proteins with affinity for metal ions are retained on a column that contains immobilised metal ions (such as cobalt, nickel, or copper) for the purification of histidine-containing peptides or proteins. Because many naturally occurring proteins do not have an affinity for metal ions, recombinant DNA technology is often used to introduce a poly-histidine tail to the target gene. When loaded onto a column, proteins containing a higher number of histidine residues are then able to bind to the column more tightly than those with fewer histidine residues.

By far, the most widely-used technique for this application is to use an immobilised nickel column, and to insert poly-histidine tags of six or more residues onto the recombinant proteins of interest, at either the *C*- or *N*-terminus. The bound protein is commonly eluted using competitive elution (for example with an imidazole gradient), stripping elution (using chelating agents, such as EDTA or EGTA, to strip the metal ions from the column), or by pH adjustment.[26]

2.1.5 Ion exchange chromatography

Protein purification by ion exchange is possible due to the fact that most proteins bear non-zero net electrostatic charges at all pHs except when the pH is equal to the pI

(isoelectric point). At a pH greater than pI, the protein becomes negatively charged. Conversely, at a pH lower than pH, the protein becomes positively charged. Ion exchange chromatography is based on the electrostatic attraction between buffer-dissolved charged proteins and oppositely charged binding sites on a solid ion exchange adsorbent.

The ion exchange adsorbent usually consists of inert spherical porous beads with charged functional groups densely grafted onto the surface. Typically, a protein mixture is transferred into a low ionic strength buffer and loaded onto a column which has been pre-equilibrated with the same buffer. When the protein mixture is loaded onto the column, proteins of the opposite charge to the column are retained whilst all other proteins simply pass through the column. The retained proteins are then eluted by gradually increasing the ionic strength of the buffer, whereby the proteins are eluted in order of increasing net charge. The buffer type, the buffer concentration, the pH, and the temperature all have important roles in controlling the separation.

The enzymes in this chapter were purified by cation exchange chromatography. The polystyrene beads used in this case were functionalised with sulfonated groups (Figure 16).

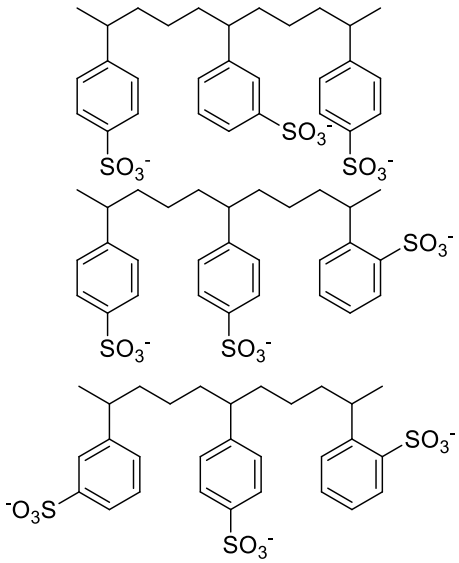


Figure 16 - Schematic of an acidic sulfonated polystyrene cation exchange resin.

The protein of interest was kept at a pH below its pI, giving the protein an overall positive charge. The protein which bound to the column was eluted using an increasing gradient of NaCl, as a competitive cationic binder.

2.1.6 Size exclusion chromatography

Size exclusion chromatography (SEC) separates macromolecules according to their hydrodynamic volume, which is defined by the Stokes radius.[27] The column consists of porous polymer beads designed to have pores of differing sizes. Particles with smaller hydrodynamic volumes equilibrate into these pores when a mobile phase is passed through the column. In this way, smaller particles have a longer path length than larger particles, and therefore molecules of different sizes become separated along the length of the column.

SEC is frequently used as a final step in protein purification and to discern between high and low molecular weight components in a mixture. It is an important means to assess the homogeneity of the sample as aggregated material can be easily separated

from a folded protein due to the molecular weight differences between free and clustered states. Additional advantages of SEC are that is operationally simple to perform and that there is no requirement for specialised equipment.

2.1.7 Circular dichroism (CD) spectroscopy

UV Circular dichroism spectroscopy is a spectroscopic technique wherein the optical rotations of molecules are measured over a range of wavelengths. Due to the chiral nature of the constituent amino acids of the peptide backbone, CD is useful for estimating the extent of secondary structure present and therefore the conformation of proteins. Since the secondary structure of a protein is sensitive to its environment, temperature and pH, CD spectroscopy can be used to observe how the secondary structure of a protein changes with environmental conditions, or upon interaction with other molecules.

Secondary structure can be determined by CD spectroscopy in the "far-UV" region of the electromagnetic spectrum (190 - 250 nm), wavelengths which encompasses the chromophore of the peptide bond[28]; a signal arises when the bond is located in a regular, folded environment. α-Helices, β-sheets, and random coil structures each give rise to a characteristic shape and magnitude in the CD spectrum.

2.2 Objectives

This chapter outlines the work done on the family GH101 enzymes, engBL[14] from *Bifidobacterium longum* and engEF[16] from *Enterococcus faecalis*, respectively. Both WT proteins have been well characterised in terms of their kinetics and substrate specificity. The expression and purification of these WT enzymes is described and the

mutagenesis of engBL and engEF, as well as their characterisation using the aforementioned techniques, is also reported.

2.3 EngBL from *Bifidobacterium longum*

2.3.1 Expression and purification of WT engBL

The plasmid encoding the protein engBL (340~1694, pET-28b) was provided by Prof. Shinya Fushinobu from the University of Tokyo, Japan. The pET-28b vector carried the sequence for an *N*-terminal His tag, and a gene which conferred kanamycin-resistance to *E. coli.*

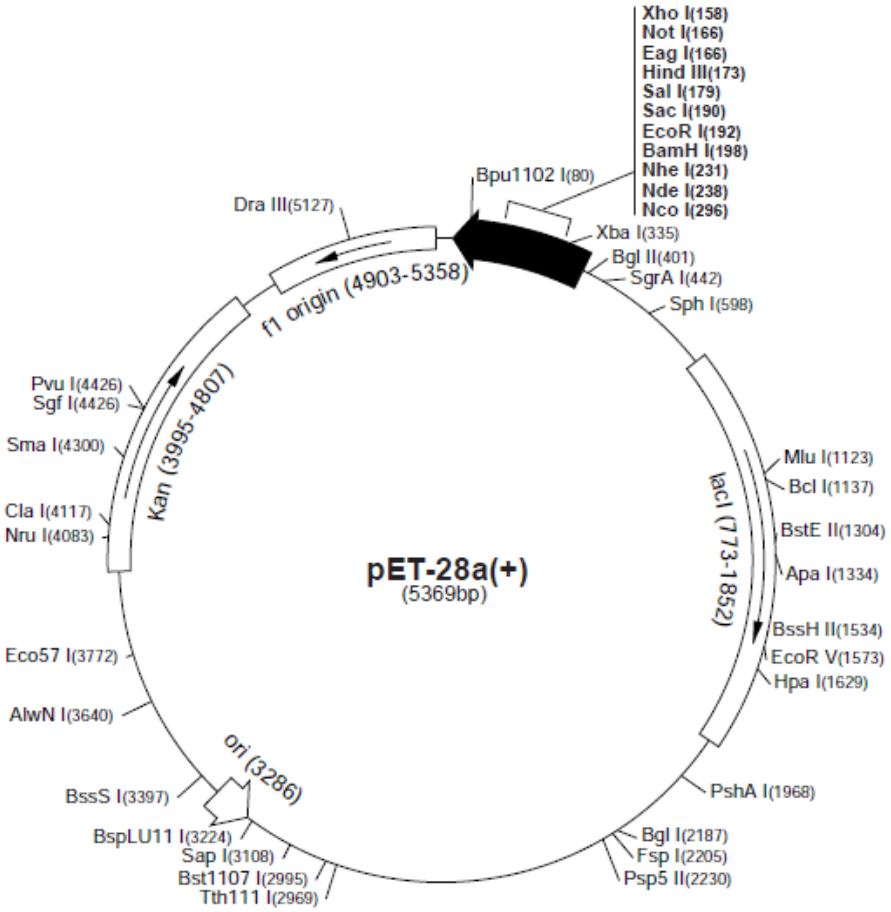


Figure 17 - The pET-28d vector. The *N*-terminal his_6-tag sequence is encoded between bases 140-157 and the kanamycin-resistance gene is encoded between 3995-4807. Adapted from Addgene vector database (https://www.addgene.org/vector-database/).

The engBL plasmid was transformed into *E. coli* (TOP10) for plasmid prep, and plasmids were isolated using a High Pure Plasmid Isolation Kit (Roche).

The engBL plasmid was then transformed into *E. coli* (BL21 DE5), grown overnight, and diluted into fresh LB media (5 mL into 600 mL), and incubated at 37 ºC. Transcription of the engBL genes were under the control of a *lac* operon, and therefore protein expression was induced by the addition of IPTG to a final concentration of 0.5 mM when cultures were at a mid-logarithmic phase (OD_{600} of 0.4 to 0.6 AU). The cells were grown overnight at 23 ˚C and then harvested by centrifugation. The pellet was resuspended in Buffer A (Table 4, Entry 1) and cells were lysed by sonication. The lysate was centrifuged at 15,000 rpm, 4 ºC for 35 min to remove the cell debris. The supernatant containing the soluble protein was filtered (0.22 µm), and applied to the affinity columns for IMAC (See Chapter 6 for detailed experimental procedures).

Adapting previously established protocols[16], a HisTrap column (GE Healthcare) pre-charged with Ni^{2+} was used to purify engBL. The clarified cell lysate was directly loaded onto the column, washed with Buffer A (Table 4, Entry 1), and eluted with an increasing gradient of Buffer B (Table 4, Entry 1).

Entry	Process	Buffer	Composition
1	Immobilised metal affinity chromatography (IMAC)	Buffer A	10 mM Tris/HCl, pH 7.5, 150 mM NaCl, 20 mM imidazole
		Buffer B	10 mM Tris/HCl, pH 7.5, 150 mM NaCl, 200 mM imidazole
2	Cation exchange chromatography (CEC)	Buffer A	10 mM Tris/HCl, pH 7.5, 20 mM NaCl
		Buffer B	10 mM Tris/HCl, pH 7.5, 1 M NaCl
3	Size exclusion chromatography (SEC)	Buffer A	10 mM Tris, 150 mM NaCl, pH 7

Table 4 - Buffers used in the purification of GH101 endo-α-*N*-acetylgalactosaminidases.

Fractions indicated by the chromatograph to contain protein were pooled and dialysed against Buffer A (Table 4, Entry 2) and loaded onto an SEC column. The protein was eluted with an increasing gradient of Buffer B (Table 4, Entry 2). Fractions indicated by the chromatograph to contain protein were pooled and concentrated to 200 µL, directly loaded onto an SEC column, and eluted with Buffer A (Table 4, Entry 3). The concentration of pooled protein after SEC was determined to be 3.5 mg/mL by means of a Nanodrop ND-1000 spectrophotometer.

The SDS-PAGE gel showed a large band at ~150 kDa corresponding to the desired protein (Figure 18).

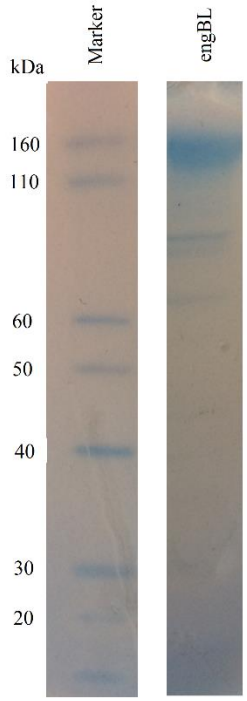


Figure 18 - SDS-PAGE analysis of engBL after SEC purification.

It was necessary then to determine the activity of WT engBL using Galβ1,3GalNAc-α-pNP (core 1-pNP) as the substrate. The International Union of Biochemistry (IUB) defines a standard unit (U) of enzyme activity as the amount of enzyme which catalyses the transformation of 1 μmol of the substrate per minute under standard conditions.[29]

A standard curve for pNP was plotted for the purposes of monitoring the hydrolysis of Galβ1,3GalNAc-α-pNP by UV spectroscopy. 10 μmol of Galβ1,3GalNAc-α-pNP was to be used in the subsequent hydrolytic assay[16], therefore the absorbance of pNP was recorded at varying quantities up to 10 μmol (Figure 19).

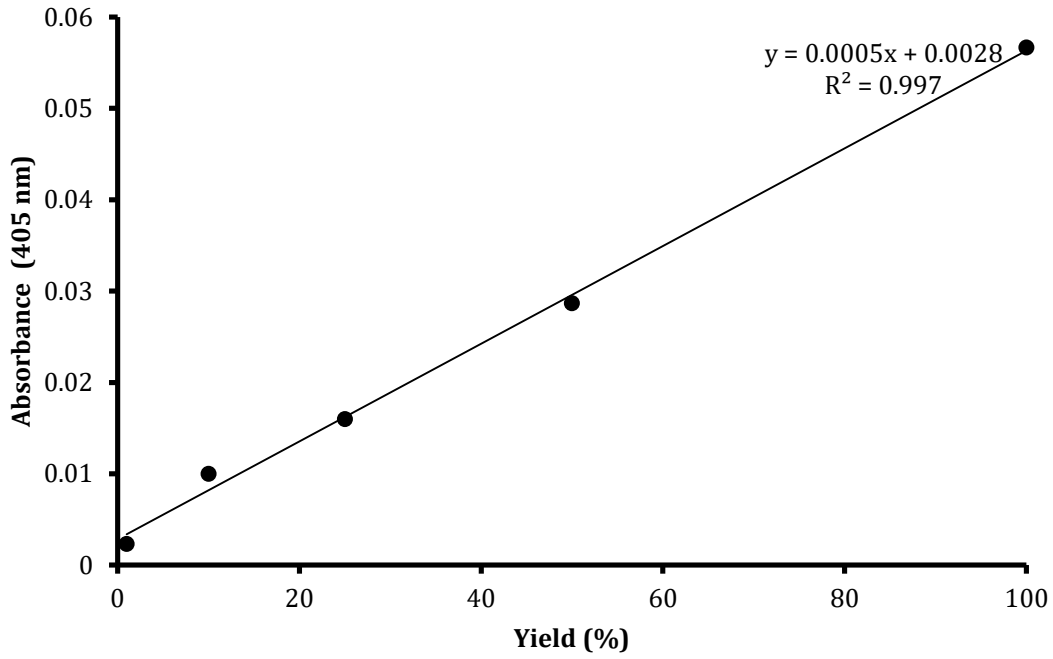
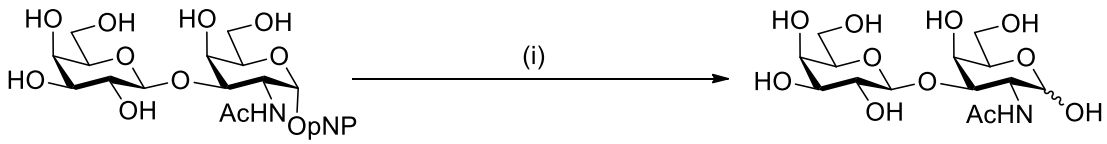


Figure 19 - Standard curve of varying pNP concentrations. The values of concentration have been converted to yield.

Galβ1,3GalNAc-α-pNP (10 µmol) was then incubated with varying dilutions (2-, 5-, 10-, 100-, 500-, and 1000-fold dilutions) of engBL for 30 mins, after which time, the reaction mixture was quenched by the addition of 1 M NaOH, and the sample analysed by UV spectroscopy at 405 nm (Scheme 17).



Scheme 17 - Determining the activity of WT engBL. Conditions: i) engBL (2 µL of various dilutions as described in the text), 25 mM NaOAc buffer, pH 6.0, 37 ˚C, 30 mins.

Whilst the reactions containing 2-, 5-, and 10-fold dilutions of engBL were complete after 30 mins, the addition of 100-fold dilution (final concentration of 1.17×10^{-7} mol/L

of enzyme) gave an absorbance of 0.037, which was sufficient for the calculation of the enzymatic activity of engBL.

Thus, substitution of 0.037 into the equation given by the standard curve in Figure 19 gives the yield of the reaction after 30 mins:

$$0.037 = 0.0005x + 0.0028$$

x = 68.4 % or 6.84 μmol of pNP released after 30 mins

The amount of pNP hydrolysed per minute is therefore:

$$\frac{6.84\ \mu\text{mol}}{30\ \text{mins}} = 0.228\ \mu\text{mol/min}$$

If 0.07 μg of engBL catalyses the hydrolysis of Galβ1,3GalNAc-α-pNP (core 1-pNP) at a rate of 0.228 μmol/min, then the amount of enzyme required to perform the hydrolytic reaction at a rate of 1 μmol/min is:

$$1\ \text{U} = \frac{0.07\ \mu\text{g}}{0.228} = 0.31\ \mu\text{g}$$

These results also confirmed that WT engBL was active as described.[13]

2.3.2 Mutagenesis of WT engBL

The primary sequence of engBF has previously been published,[30] and more recently, the crystal structure of WT engBF has also been reported.[15] Mutational analyses revealed the identity of the important residues for substrate binding, with two Trp residues (W748 and W750) forming stacking interactions with the β-faces of the sugar rings of Galβ1,3GalNAc by substrate-induced fit. The two key catalytic residues were determined to be D789 and E822, which act as the catalytic nucleophile and catalytic general acid/base, respectively.

We expected that the mutation of the nucleophilic residue should lead to the abolition of the hydrolytic activity of engBL.[9] Therefore, Asp789 was substituted by site-directed mutagenesis to generate glycosynthases. Three common mutations that are normally conferred to generate glycosynthases involve converting the nucleophile to glycine, alanine, and serine.[18,31] These three residues are non-nucleophilic (in the cases of glycine and alanine), and smaller in size. The codon of the engBL gene corresponding to D789 was identified using ClustalW[12]. Primers were designed for site-directed mutagenesis by replacing the codon corresponding to D789 with the codons corresponding to the desired mutation (Table 5) as described in Section 2.1.2.[18,19]

Mutation	Primer	Sequence (5' - 3')
D789G	Fwd	GAC TTC ATC TAC CTC **GGG** GTG TGG GGC AAC CTG
	Rev	CAG GTT GCC CCA CAC **CCC** GAG GTA GAT GAA GTC
D789A	Fwd	GAC TTC ATC TAC CTC **GCT** GTG TGG GGC AAC CTG
	Rev	CAG GTT GCC CCA CAC **AGC** GAG GTA GAT GAA GTC
D789S	Fwd	GAC TTC ATC TAC CTC **AGC** GTG TGG GGC AAC CTG
	Rev	CAG GTT GCC CCA CAC **GCT** GAG GTA GAT GAA GTC

Table 5 - Primers for the mutagenesis of engBL. Codons corresponding to D789 are in bold.

The site directed mutagenesis was performed using the Quikchange® II Site-Directed Mutagenesis Kit (Agilent) with Phusion® High-Fidelity DNA Polymerase (Thermo Scientific), as described in Chapter 6, on the WT plasmid in the presence of the respective primers. The parent plasmid was then digested with *DpnI*, and the mutagenesis products were directly transformed into *E. coli* (DH5α) for plasmid prep and grown overnight. The plasmids were sequenced to confirm the desired mutation had been incorporated. The mutated products were then transformed and expressed in *E. coli* (BL21 DE5), and purified as described earlier for the WT engBL enzyme.

SDS-PAGE analysis showed the presence of a protein at ~150 kDa corresponding to the mutants of engBL (Figure 20).

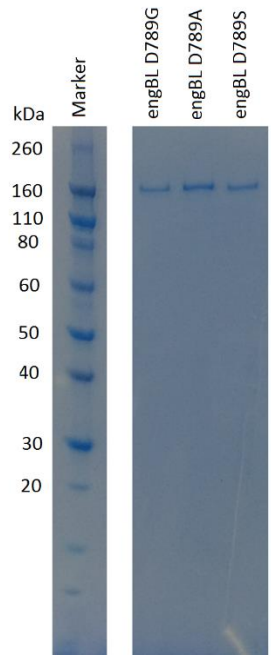


Figure 20 - SDS-PAGE Analysis of engBL mutants after purification.

The CD spectra of WT and variant engBL proteins were compared and were found to be similar (Figure 21). These results indicate that displacement of the nucleophile Asp789 had no effects on protein folding.

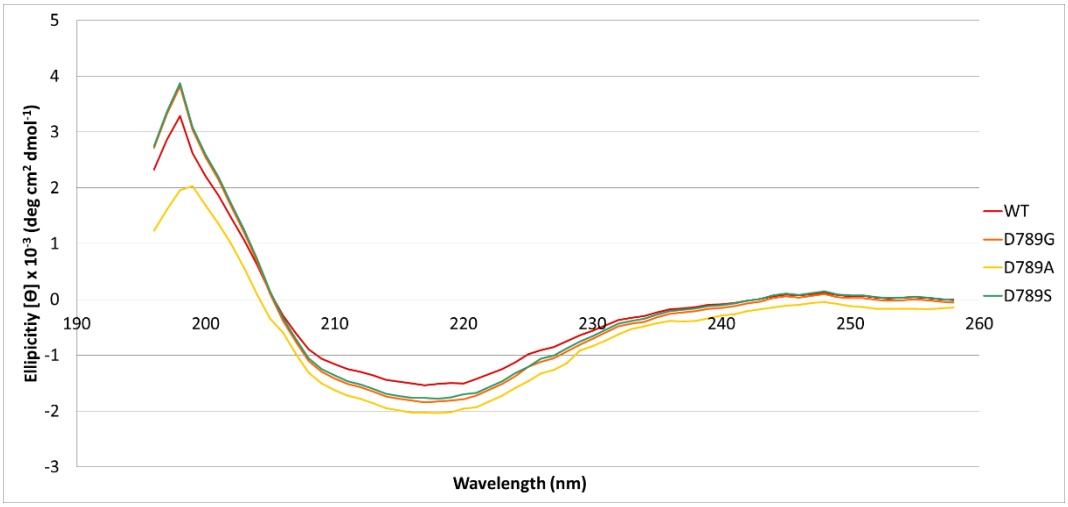


Figure 21 - CD spectrum of WT engBL and the D789 mutants.

The concentrations of pooled protein for engBL D789G, D789A, and D789S after SEC were determined to be 9.5, 3.6, and 12.2 mg/mL respectively by means of a Nanodrop ND-1000 spectrophotometer.

2.4 EngEF from *Enterococcus faecalis*

2.4.1 Expression and purification of WT engEF

The plasmid for engEF (pET-23b) was obtained from Dr. Makoto Ito of Kyushu University, Japan. The pET-23b vector contained both an *N*-terminal and *C*-terminal his tag. However only the *C*-terminal his tag is expressed in this construct (Figure 22). The vector also had a gene for ampicillin-resistance.

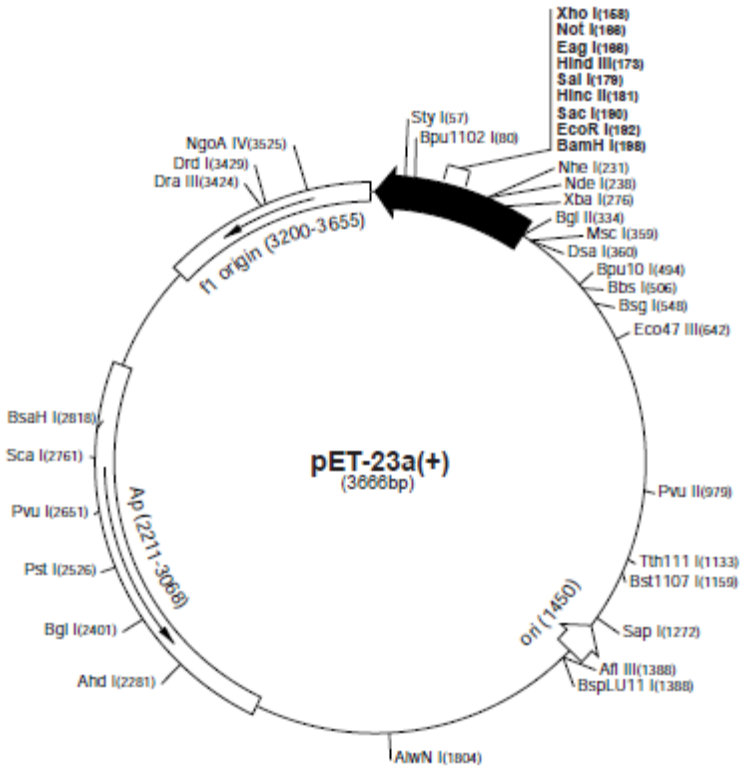


Figure 22 - The pET-23b vector. Adapted from Addgene vector database (https://www.addgene.org/vector-database/)

The plasmid was transformed into *E. coli* (TOP10) for plasmid prep, and plasmids were isolated using a High Pure Plasmid Isolation Kit (Roche).

Several *E. coli* cell lines were tested to optimise the expression of engEF, namely BL-21*, Chaperone 3, and codon plus cell lines. The Chaperone 3 cell line contains genes encoding chaperone proteins which, when expressed, assist with the correct folding of the target protein, whilst the codon plus cell line includes extra codons that are not native to *E. coli*, but may be required for protein expression. SDS-PAGE analysis indicated that the Chaperone 3 cell line expressed a greater amount of protein compared to either BL-21* or the codon plus cell lines (gel not shown). Therefore, engEF was transformed into *E. coli* (Chaperone 3), and grown in LB media (1 L). Transcription of the engEF gene was under the control of a *lac* operon, and therefore protein expression was induced by the addition of IPTG to a final concentration of 0.5 mM when cultures were at a mid-logarithmic phase (OD_{600} of 0.4 to 0.6 AU), grown overnight at 23 ˚C, and harvested by centrifugation. The pellet was resuspended in Buffer A (Table 4, Entry 1) and the cells lysed by sonication. The lysate was centrifuged at 15,000 rpm, 4 ˚C for 35 min to remove the cell debris. The supernatant containing the soluble protein was filtered (0.22 μm) and applied to the affinity columns.

Adapting protocols established previously,[16] a HisTrap column (GE Healthcare) pre-charged with Ni^{2+} was used to purify engEF. The clarified cell lysate was directly loaded onto the column, washed with elution buffer (Buffer A, Table 4), and eluted with an increasing gradient of Buffer B (Table 4, Entry 1). Fractions indicated by the chromatograph to contain protein were analysed by SDS-PAGE. The SDS-PAGE gel showed a large band at ~150 kDa corresponding to the desired protein (Figure 23).

The concentration of pooled protein was determined to be 1.5 mg/mL by means of a Nanodrop ND-1000 spectrophotometer.

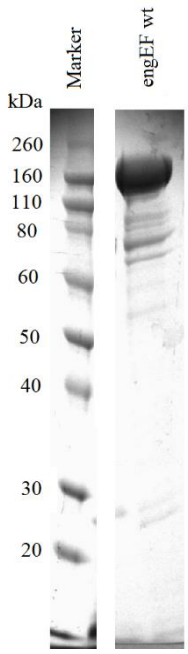
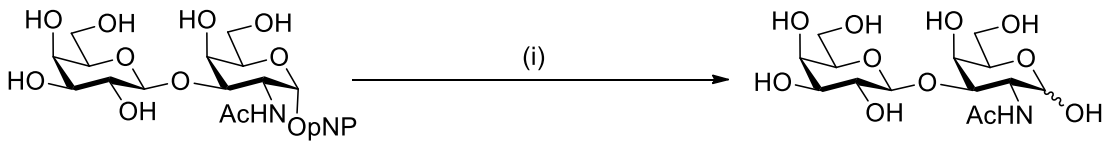


Figure 23 - SDS-PAGE analysis of WT engEF after purification by IMAC.

As with WT engBL, the activity of engEF was determined. Galβ1,3GalNAc-α-pNP (10 µmol) was incubated with varying dilutions (2-, 5-, 10-, 100-, 500-, and 1000-fold dilutions) of engBL for 30 mins, after which time, the reaction mixture was quenched by the addition of 1 M NaOH, and the sample analysed by UV spectroscopy at 405 nm (Scheme 18).



Scheme 18 - Testing the activity of WT engEF. Conditions: i) engEF (2 µL of various dilutions as described in text), 25 mM NaOAc buffer, pH 6.0, 37 ˚C, 30 mins.

Whilst the reactions containing 2-, 5-, and 10-fold dilutions of engEF were complete after 30 mins, the addition of 100-fold dilution (final concentration of 1.02 x 10^{-7} mol/L of enzyme) gave an absorbance of 0.019, which was sufficient for the calculation of activity.

Thus, substitution of 0.019 into the equation given by the standard curve in Figure 19 gives the yield of the reaction after 30 mins:

$$0.019 = 0.0005x + 0.0028$$

x = 32.4 % or 3.24 µmol of pNP released after 30 mins

The amount of pNP hydrolysed per minute is:

$$\frac{3.24\ \mu\text{mol}}{30\ \text{mins}} = 0.108\ \mu\text{mol/min}$$

Therefore, if 0.03 µg of engEF catalyses the hydrolysis of Galβ1,3GalNAc-α-pNP (core 1-pNP) at a rate of 0.108 µmol/min, then the amount of enzyme required to perform the hydrolytic reaction at a rate of 1 µmol/min is:

$$1\ \text{U} = \frac{0.03\ \mu\text{g}}{0.108} = 0.27\ \mu\text{g}$$

These results also confirmed that WT engEF was active as described.[16]

2.4.2 Mutagenesis of engEF

Two putative catalytic aspartic acid residues were identified in engBL to be D628 and D789.[13] A sequence alignment of engBL and engEF using ClustalW[12] indicated that the amino acids corresponding to the nucleophilic and catalytic acid/base residues were D725 and E754, respectively (Figure 24).

```
engBL   745 MVRRNSAGGLSYGWNWLDQGVGIDGIYDLASGS 767
engEF   677 LVDP-----TKRGWDWLDPSYFIKQRPDALSGR 704

engBL   768 RVSRFADLSKEVGDNMDFIYLDVWGNLTSSGSE 800
engEF   705 RYERFKELKQ-KAPNLDYIYVDVWGNQGE---- 732

engBL   801 DSWETRKMSKMINDNGWRMTTEWGSGNEYDSTF 833
engEF   733 SGWASRQLSKEINSLGWFTTNEFPNALEYDSVW 765
```

Figure 24 - Sequence alignment of engBL and engEF for the identification of the catalytic nucleophile (yellow) and catalytic acid/base (turquoise) in engEF.

Replacement of the catalytic nucleophile (D725) with non-nucleophilic residues should remove all hydrolytic activity of engEF. The codon of the engEF gene corresponding to D725 was identified using ClustalW[12]. Primers were designed for site-directed mutagenesis by replacing the codon corresponding to D725 with the codons corresponding to the desired mutations (Table 6) as described in Section 2.1.2.[18,19]

Mutation	Primer	Sequence (5' - 3')
D725G	Fwd	GAT TAT ATT TAT GTT **GGG** GTT TGG GGC AAT CAA GCC
	Rev	GGC TTG ATT GCC CCA AAC **CCC** AAC ATA AAT ATA ATC
D725A	Fwd	GAT TAT ATT TAT GTT **GCT** GTT TGG GGC AAT CAA GCC
	Rev	GGC TTG ATT GCC CCA AAC **AGC** AAC ATA AAT ATA ATC
D725S	Fwd	GAT TAT ATT TAT GTT **AGC** GTT TGG GGC AAT CAA GCC
	Rev	GGC TTG ATT GCC CCA AAC **GCT** AAC ATA AAT ATA ATC

Table 6 - Primer design for the mutagenesis of engEF. Codons corresponding to D725 are in bold.

Site directed mutagenesis was performed using a Quikchange® II Site-Directed Mutagenesis Kit (Agilent) with Phusion® High-Fidelity DNA Polymerase (Thermo Scientific), as described in Chapter 6, using the WT plasmid in the presence of the respective primers. The parent plasmid was then digested with *DpnI*, and the mutagenesis products were directly transformed into *E. coli* (DH5α) for plasmid prep, and grown overnight. The plasmids were sequenced to confirm the desired mutations had been incorporated. The mutated products were transformed and expressed in *E. coli* (BL21 DE5), and then purified as described earlier for the WT engBL enzyme (Section 2.3.1).

The SDS-PAGE gel showed a large band at ~150 kDa corresponding to the desired protein (Figure 25).

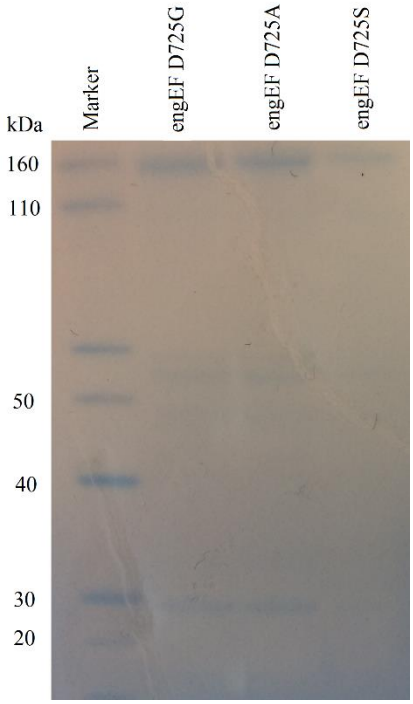


Figure 25 - SDS-PAGE analysis of engEF mutants.

The CD spectra of WT and variant engEF were compared and were found to be similar (Figure 26). These results indicated that displacement of the nucleophile Asp725 had no effect on protein folding.

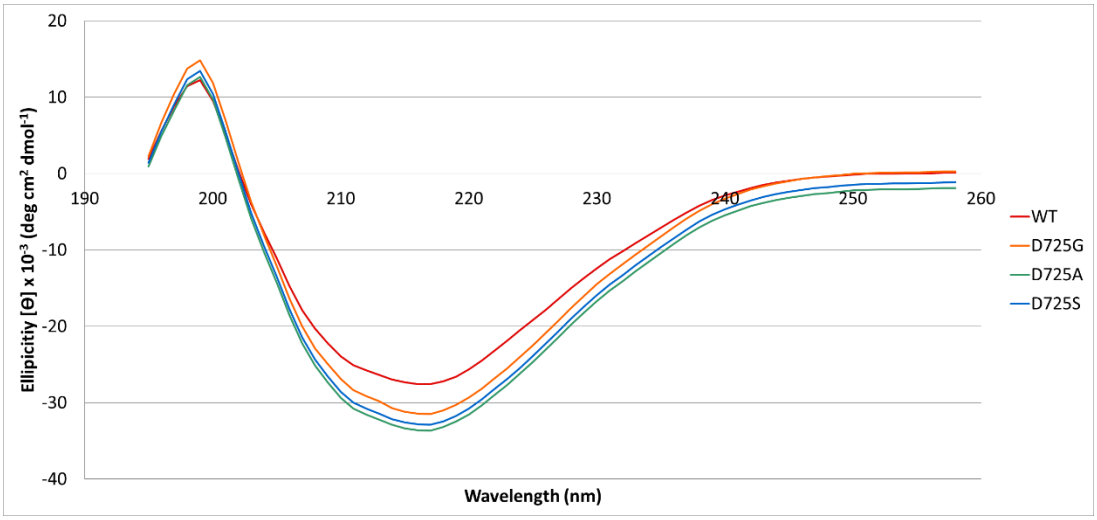


Figure 26 - CD spectrum of engEF and the D725 mutants.

The concentration of pooled protein for engEF D725G, D725A, and D725S after SEC were determined to be 1.4, 9.4, and 8.6 mg/mL respectively by means of a Nanodrop ND-1000 spectrophotometer.

2.5 Conclusions

To date, there has been no published data on the engineering of α-glycosynthases from any retaining GH101 endo-α-*N*-acetylgalactosaminidases. The work described in this chapter illustrates an attempt to engineer glycosynthase enzymes from two of the members of family GH101: engBL from *Bifidobacterium longum* and engEF from *Enterococcus faecalis*.

The WT enzymes, engBL and engEF, from *Bifidobacterium longum and Enterococcus faecalis* respectively, were successfully transformed into *E. coli* (BL21 DE5) competent cells. Subsequent protein expression *via* ITPG induction, followed by purification of the WT enzymes were successful. EngBL was isolated at a concentration of 3.5 mg/mL, whilst engEF was isolated at a concentration of 1.5 mg/mL.

Site directed mutagenesis was performed on the genes of engBL and engEF where the catalytic nucleophilic residue of the respective glycosidases was substituted for individual glycine, alanine, and serine mutants. DNA sequencing confirmed the substitutions of the appropriate amino acid for the respective proteins.

Plasmid transformation, protein expression and purification of the engBL D789G/A/S and engEF D725G/A/S enzymes were all successful. Analysis of the mutant

glycosidases by CD spectroscopy indicated that the proteins were correctly folded. The kinetic activity of these proteins was investigated and is reported in Chapter 4.

2.6 References

(1) Schmaltz, R. M.; Hanson, S. R.; Wong, C. -H. *Chem. Rev.* **2011**, *111* (7), 4259–4307.

(2) Crout, D. H.; Vic, G. *Curr. Opin. Chem. Biol.* **1998**, *2* (1), 98–111.

(3) Koeller, K. M.; Wong, C.-H. *Chem. Rev.* **2000**, *100* (12), 4465–4494.

(4) Wang, L. X.; Huang, W. *Curr. Opin. Chem. Biol.* **2009**, *13* (5-6), 592–600.

(5) Cobucci-Ponzano, B.; Trincone, A.; Giordano, A.; Rossi, M.; Moracci, M. *Biochemistry* **2003**, *42* (32), 9525–9531.

(6) Cobucci-Ponzano, B.; Conte, F.; Mazzone, M.; Bedini, E.; Corsaro, M. M.; Rossi, M.; Moracci, M. *Biocatal. Biotransfor.* **2008**, *26* (1-2), 18–24.

(7) Usui, T.; Matsui, H.; Isobe, K. *Carbohydr. Res.* **1990**, *203* (1), 65–77.

(8) Mackenzie, L. F.; Wang, Q.; Warren, R. A. J.; Withers, S. G. *J. Am. Chem. Soc.* **1998**, *120* (22), 5583–5584.

(9) Cobucci-Ponzano, B.; Strazzulli, A.; Rossi, M.; Moracci, M. *Adv. Synth. Catal.* **2011**, *353* (13), 2284–2300.

(10) CAZy Database (Carbohydrate-Active enZYmes) http://www.cazy.org/.

(11) Koshland, D. E. *Biol. Rev.* **1953**, *28* (4), 416–436.

(12) Thompson, J. D.; Gibson, T. J.; Higgins, D. G. In *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2002.

(13) Fujita, K.; Oura, F.; Nagamine, N.; Katayama, T.; Hiratake, J.; Sakata, K.; Kumagai, H.; Yamamoto, K. *J. Biol. Chem.* **2005**, *280* (45), 37415–37422.

(14) Ashida, H.; Ozawa, H.; Fujita, K.; Suzuki, S.; Yamamoto, K. *Glycoconj. J.* **2010**, *27* (1), 125–132.

(15) Suzuki, R.; Katayama, T.; Kitaoka, M.; Kumagai, H.; Wakagi, T.; Shoun, H.; Ashida, H.; Yamamoto, K.; Fushinobu, S. *J. Biochem.* **2009**, *146* (3), 389–398.

(16) Goda, H. M.; Ushigusa, K.; Ito, H.; Okino, N.; Narimatsu, H.; Ito, M. *Biochem. Biophys. Res. Commun.* **2008**, *375* (4), 541–546.

(17) Koutsioulis, D.; Landry, D.; Guthrie, E. P. *Glycobiology* **2008**, *18* (10), 799–805.

(18) Pozzo, T.; Plaza, M.; Romero-García, J.; Faijes, M.; Karlsson, E. N.; Planas, A. *J. Mol. Catal. B Enzym.* **2014**, *107*, 132–139.

(19) Cobucci-Ponzano, B.; Zorzetti, C.; Strazzulli, A.; Bedini, E.; Corsaro, M. M.; Sulzenbacher, G.; Rossi, M.; Moracci, M. *Biocatal. Biotransfor.* **2012**, *30* (3), 288–295.

(20) Ling, M. M.; Robinson, B. H. *Anal. Biochem.* **1997**, *254* (2), 157–178.

(21) Zheng, L.; Baumann, U.; Reymond, J.-L. *Nucleic Acids Res.* **2004**, *32* (14), e115–e115.

(22) Kirsch, R. D.; Joly, E. *Nucleic Acids Res.* **1998**, *26* (7), 1848–1850.

(23) Edelheit, O.; Hanukoglu, A.; Hanukoglu, I. *BMC Biotechnol.* **2009**, *9* (1), 61.

(24) Lundberg, K. S.; Shoemaker, D. D.; Adams, M. W. W.; Short, J. M.; Sorge, J. A.; Mathur, E. J. *Gene* **1991**, *108* (1), 1–6.

(25) Mierzejewska, K.; Siwek, W.; Czapinska, H.; Kaus-Drobek, M.; Radlinska, M.; Skowronek, K.; Bujnicki, J. M.; Dadlez, M.; Bochtler, M. *Nucleic Acids Res.* **2014**, *42* (13), 8745–8754.

(26) Paunovic, I.; Schulin, R.; Nowack, B. *J. Chromatogr. A* **2005**, *1100* (2), 176–184.

(27) Erickson, H. P. *Biol. Proced. Online* **2009**, *11* (1), 32–51.

(28) Kelly, S. M.; Jess, T. J.; Price, N. C. *Biochim. Biophys. Acta - Prot. Proteom.* **2005**, *1751* (2), 119–139.

(29) International Union of Biochemistry (1965) *Enzyme Nomenclature, Recommendations 1964 of the International Union of Biochemistry*, Elsevier, Amsterdam.

(30) Fukuda, S.; Toh, H.; Hase, K.; Oshima, K.; Nakanishi, Y.; Yoshimura, K.; Tobe, T.; Clarke, J. M.; Topping, D. L.; Suzuki, T.; Taylor, T. D.; Itoh, K.; Kikuchi, J.; Morita, H.; Hattori, M.; Ohno, H. *Nature* **2011**, *469* (7331), 543–547.

(31) Cobucci-Ponzano, B.; Conte, F.; Bedini, E.; Corsaro, M. M.; Parrilli, M.; Sulzenbacher, G.; Lipski, A.; Dal Piaz, F.; Lepore, L.; Rossi, M.; Moracci, M. *Chem. Biol.* **2009**, *16* (10), 1097–1108.

Chapter 3 – Synthesis of Activated *O*-Glycan Core Structures

3.1 Introduction

Oligosaccharides have an undeniable importance in biological systems, especially the motifs that decorate glycoproteins, the common glycone portions of a variety of antibiotics, and the intricate carbohydrate assemblies that are required for signalling processes. In fact, in a large number of cases, the specificity and activity of these biological processes rely on the carbohydrate portion of the molecule.[1]

As described in Chapter 1, the biosynthetic attachment of carbohydrates to proteins is an important post-translational process. However, the biosynthesis of glycans is not under direct genetic control, so glycoproteins are produced intracellularly as inseparable heterogenous mixtures of glycoforms, in which different oligosaccharide structures are linked to the same peptide chain. Accessing pure single glycoforms of glycoproteins has therefore become an important scientific objective in order to prepare therapeutic glycoproteins.

Conventional chemical synthetic approaches are, in most cases, unable to provide large amounts of oligosaccharides in a timely fashion due to the complexities involved in protecting group manipulations of various glycosyl donors and acceptors. To provide useful quantities of these materials, techniques that combine chemical and enzymatic methods are increasingly being used.[2]

An excellent approach to the stereoselective synthesis of glycosidic linkages involves the use of retaining glycosidases in glycosylation reactions.[3–5] According to the

established mechanism of retaining glycosidases,[6] normal hydrolysis occurs when the glycosyl intermediate of a donor sugar reacts with water (Figure 27). If the glycosyl enzyme intermediate is intercepted with a sugar, then the result will be a glycosyl transfer to afford a new glycosidic linkage with net retention of stereochemistry.

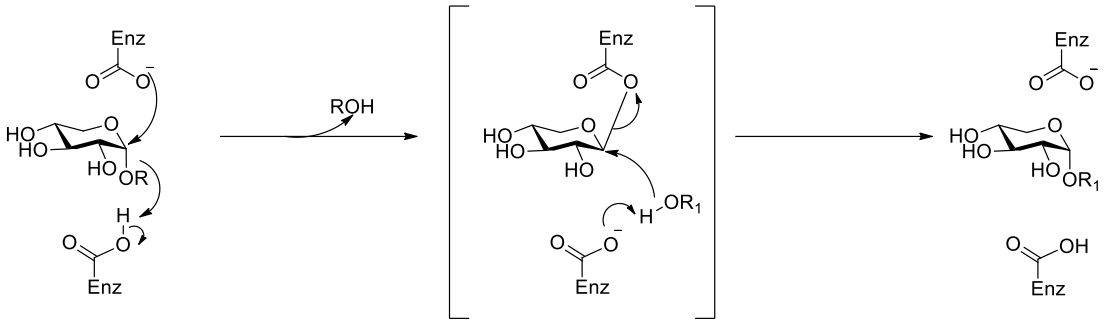


Figure 27 - Mechanism of transglycosylation with a retaining α-glycosidase.

Although there have been many reports of the use of retaining α-glycosidases as biological tools for enzymatic syntheses,[7] reports of α-glycosynthases developed from retaining α-glycosidases are scarce. Correspondingly, little synthetic work on the β-glycosyl donors that would be the required donors for these synthases has been reported. In fact, most glycosynthase enzymes to date have been engineered from retaining β-glycosidases.[8,9] The research in this thesis aims to investigate the use of α-retaining glycosidases for the synthesis of *O*-linked glycans. The rationale behind enzymatic donor design is discussed in Section 3.3.

3.1.1. Glycosyl donors for glycosidases and glycosynthases

3.1.1.1. Glycosyl fluorides

Over the last decade, glycosyl fluorides have been used numerous times as reliable donors for enzymatic glycosylations.[10] The interest in these compounds has led to the development of methods to access these in pure anomeric forms.[11–14] The small size of

fluorine, comparable to that of a hydroxyl group, means that there is minimal steric difference upon introduction of this atom to a molecule.[15]

α-Glycosyl fluorides are the only glycosyl halides stable enough to be dissolved in an aqueous buffer without appreciable rates of spontaneous hydrolysis. In fact, in an aqueous solution, most α-glycosyl fluorides are stable for periods of weeks or longer, especially if kept below 0 ºC and buffered to minimize auto-decomposition arising from acid-catalysed solvolysis.[10] In contrast, β-glycosyl fluorides are less stable, but are still useful substrates provided that the rates of enzymatic reactions are sufficiently high. Measurements made of the spontaneous rates of hydrolysis of α- and β-D-glucopyranosyl fluoride in 1 M $NaClO_4$ at 50 ºC showed that the β-fluoride hydrolyses around 40-fold faster than the α-fluoride.[16]

However, 2-acetamido-2-deoxy glycosyl fluorides with a 1,2-*trans* configuration hydrolyse rapidly in water (7.1×10^{-3} sec^{-1} at 25 ˚C) or methanol due to neighbouring group participation[17], which makes the synthesis of such donors very difficult.

Since their first reported use, glycosyl fluorides have been proven particularly useful as substrates for glycosidases. Their employment as substrates has been associated with high substrate turnover rates, which make them very useful as substrates for determining the stereochemical outcome of enzyme-catalysed glycoside hydrolysis since the hydrolysis reaction is much faster than mutarotation.[18,19]

Before the development of glycosynthases, glycosyl fluorides had been used as substrates for enzymatic oligosaccharide synthesis *via* the Hehre re-synthesis mechanism.[20] In this process, an inverting glycosidase is presented with an activated

substrate possessing the ‘wrong’ anomeric configuration (i.e. the opposite configuration to that of the natural substrate). The hydrolysis reaction is believed to proceed by way of an initial transglycosylation from one glycosyl fluoride molecule to another to form a glycosidic linkage with inversion of anomeric stereochemistry (Figure 28). The transglycosylation product, which has the ‘correct’ anomeric stereochemistry for the enzyme, is then rapidly hydrolysed, affording the product with inverted stereochemistry. The net result is retention of stereochemistry from the initial glycosyl fluoride. In some cases, the intermediate transglycosylation products can be isolated.[21–24]

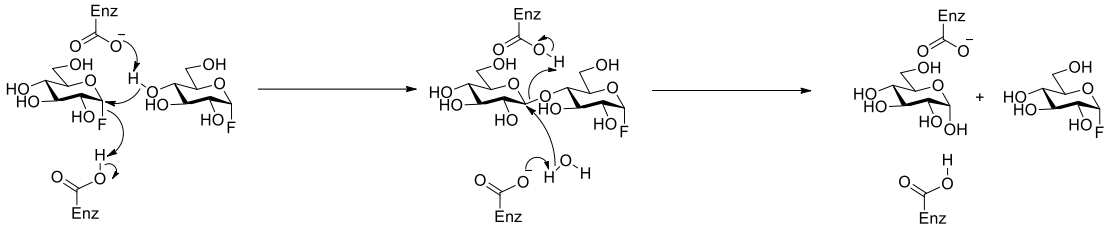


Figure 28: Hehre resynthesis-hydrolysis mechanism for the hydrolysis of the 'wrong' anomeric fluoride using inverting β-glycosidases.

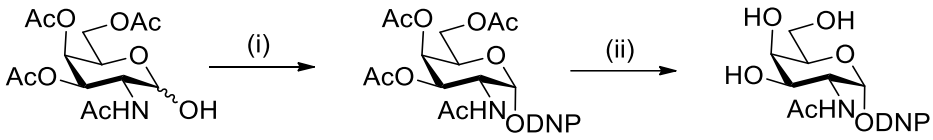
Glycosyl fluorides have been successfully used in multiple glycosidase-related studies including as probes for enzyme mechanism,[25,26] for enzyme inhibition,[27,28] and for investigations of kinetic isotope effects.[29,30] Their uses as donors for glycosidases,[22,31–38] donors for glycosyl transferases,[39] and donors for glycosynthase reactions[40] has also been well established. As stated before, β-glycosyl fluorides containing a 2-acetamido functionality with a 1,2-*trans* configuration are readily hydrolysed in aqueous solution. Investigations into the synthesis of β-glycosyl fluorides are described in Chapter 4, and an attempt was made following a literature precedent to synthesise β-fluorides of 2-

acetamido sugars (see Section 5.3.1). However, they were later discounted as viable substrates for the enzymatic studies of this thesis due to their lability.

3.1.1.2. Dinitrophenol glycosides

Originally introduced by Capon and Thomson,[41] 2,4-dinitrophenyl (2,4-DNP) glycosides are one of the most useful classes of chromogenic substrates used to assay glycosidase activity.[42] The utility of DNP glycosides arises due to the low pK_a of the 2,4-DNP leaving group (~4.0).[43–45] Ballardie *et al*. have determined the rate constant for the spontaneous hydrolysis of 2,4-dinitrophenyl β-D-glucopyranoside to be 1 x 10^5 sec^{-1} at 25 ºC, 10-20 times faster than that of the corresponding glucoside.[17] Due to the low pK_a of 2,4-dinitrophenol, performance of continuous assays at relatively low pH values is possible, which is important in some cases, for example for the kinetic analyses of lysosomal glycosidases whose optimal pH is typically around pH 4-5.

Direct glycosylation of dinitrophenol to sugars using conventional glycosylation protocols has proven to be difficult, especially in the cases of sugars containing an *N*-acetamido functionality at the 2-position.[42] van Boom *et al*. reported an alternative approach wherein a partially protected sugar hemiacetal was reacted with 1-fluoro-2,4-dinitrobenzene in the presence of a hindered base, typically 1,4-diazabicyclo[2.2.2]octane (DABCO).[46] In this nucleophilic aromatic substitution reaction, the kinetic β-glycoside was often obtained in good yields. Facile isomerisation to the α-glycoside can be achieved, if desired, by treatment with a stronger base, such as potassium carbonate, in solvents such as DMF or DMSO (Scheme 19).[46] The mechanism of the isomerization has been discussed by Berven *et al*.[47,48]



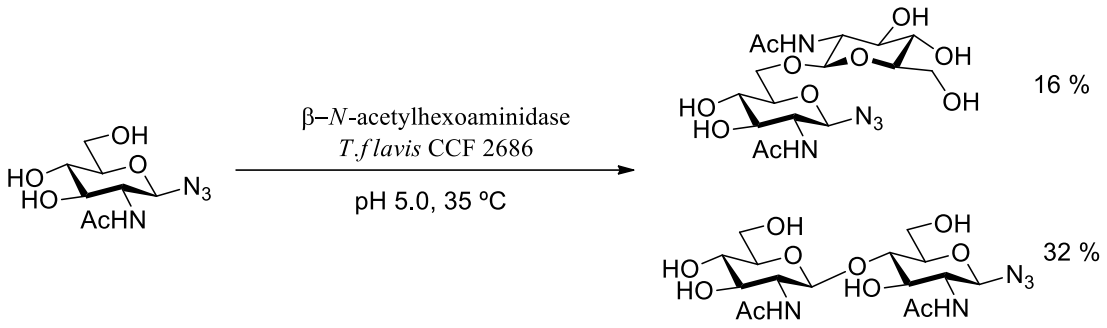
Scheme 19 - Synthesis of 2,4-dinitrophenyl 2-acetamido-2-deoxy-α-D-glucopyranoside. i) 1-fluoro-2,4-dinitrobenzene, K_2CO_3, DMF, rt; (ii) AcCl, methanol, 0 ºC.

Although there have been several methods published on the syntheses of DNP glycosides,[17,49–51] these glycosylations have rarely focussed on hexoses with a 2-acetamido functionality. In fact, Chen and Withers[51] found that the extreme lability of 2,4-dinitrophenyl 2-acetamido-2-deoxy-β-D-glucopyranoside due to neighbouring group participation of the amide made their synthetic attempts fruitless.

Attempts at the synthesis of DNP glycosides are also discussed in Chapter 5. However due to the aforementioned shortfall of using 2,4-dinitrophenyl β-*N*-acetylhexosamines as donors for glycosidases, they were not used for the work in this thesis.

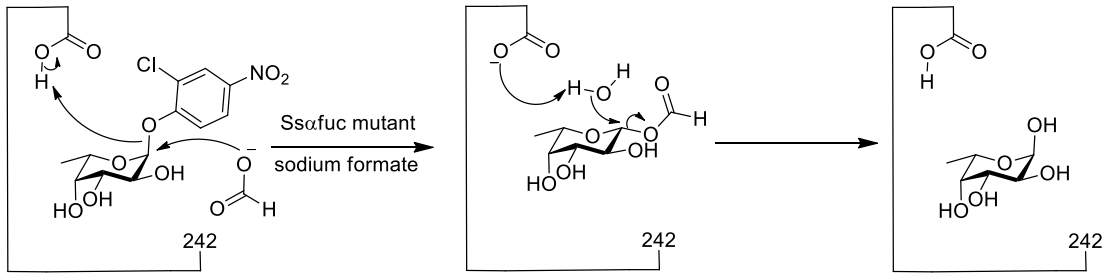
3.1.1.3. Glycosyl azides

In 2005, Fiavolá *et al*. found that the β-*N*-acetylhexosaminidase from *Talaromyces flavus* CCF2686 could catalyse the self-condensation of *N*-acetyl-β-D-glucopyranosyl azide (β-D-GlcNAc-N_3) to give the disaccharides β-D-GlcNAc-1,4-β-D-GlcNAc-N_3 and β-D-GlcNAc-1,6-β-D-GlcNAc-N_3, demonstrating that glycosyl azides could also be viable donors for glycosidase-catalysed glycosylation reactions (Scheme 20).[52]



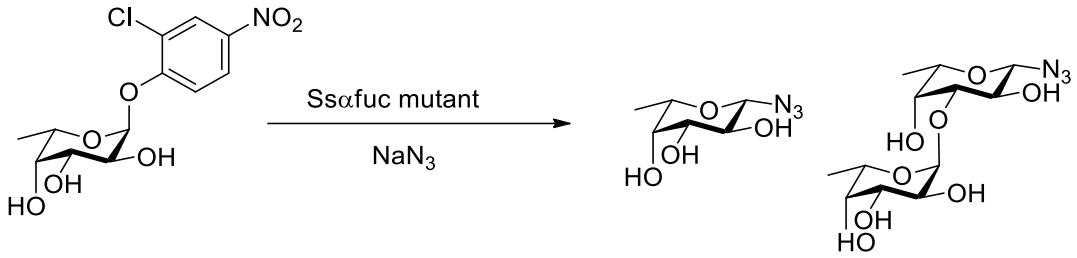
Scheme 20 - Self-condensation of β-D-GlcNAc-N_3 catalysed by β-*N*-acetylhexosaminidase from *T. flavis* CCF 2686.

Subsequently, in 2008, Cobucci-Ponzano *et al*. produced mutants of an α-L-fucosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus* (Ssαfuc, GH29).[53,54] The mutants, SsD242G/A/S, were completely inactive as expected. Whilst the initial chemical rescue experiment of the mutants with sodium formate, 4-nitrophenyl and 2-chloro-4-nitrophenyl α-L-fucopyranoside (2C4MP-Fuc) showed clear reactivation, no fucosylated oligosaccharides were produced (Scheme 21).



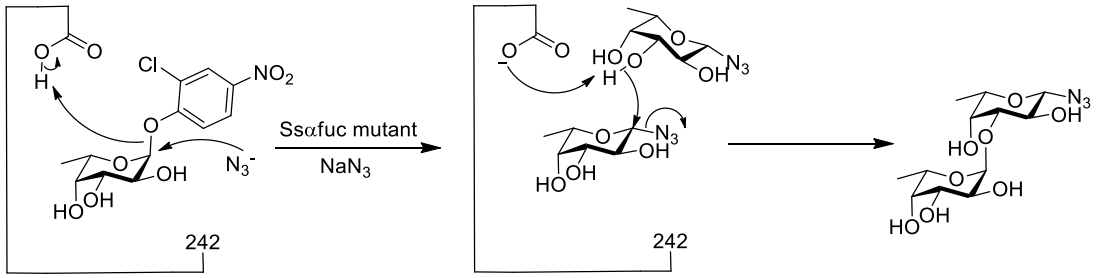
Scheme 21 – Mechanism of chemical rescue of GH29 SsD242G/A/S using sodium formate.

Interestingly, treatment of 2C4MP-Fuc with the Ssαfuc mutants in the presence of sodium azide lead to the formation of β-L-fucopyranosyl azide (β-Fuc-N_3) and the disaccharide α-L-fucopyranosyl-1,3-β-L-fucopyranosyl azide (α-L-Fuc-1,3-β-L-Fuc-N_3) (Scheme 22).



Scheme 22 - Ssαfuc mutants catalysing the self-condensation of a fucosyl azide donor to generate disaccharides.

The mechanism proposed suggested that the β-L-Fuc-N_3 product became a donor and, as a result of its accumulation, also an acceptor (Scheme 23).

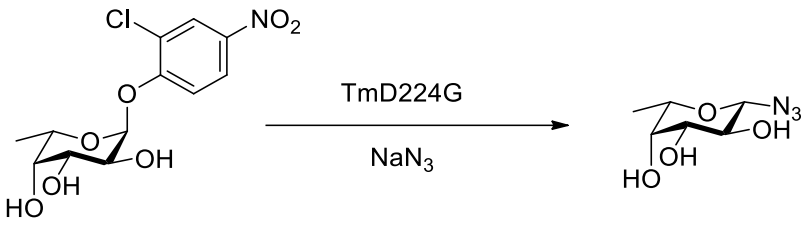


Scheme 23 - Proposed mechanism for the self-condensation of β-L-fucopyranosyl azide catalysed by the *Ssαfuc* mutants.

This result was a surprise since glycosyl azides were considered to be able compounds that were not hydrolysed by glycosidases.[55] However direct use of β-L-Fuc-N_3 as the donor with the SsD242S mutant lead to fucosylation to produce the aforementioned disaccharide in 40 % yield. In light of these results, it was then hypothesized that β-L-Fuc-N_3 could be a donor for efficient α-fucosynthases (that is, α-fucosynthases that had a high substrate turnover rate with respect to the glycosyl azide).

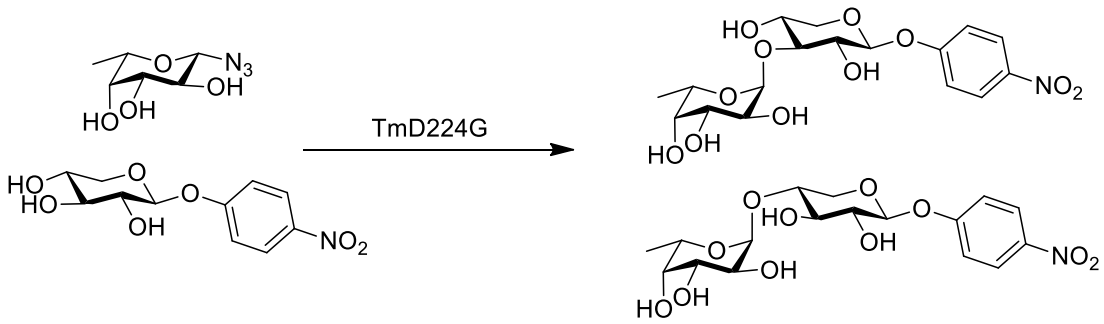
The same methodology was later applied to an α-L-fucosidase from *Thermotoga maritima* (Tmα-fuc).[56,57] The activity of a TmD224G mutant was rescued with sodium

azide using 4-nitrophenyl and 2-chloro-4-nitrophenyl α-L-fucopyranoside substrates.[58] Analysis of the reaction mixture revealed that the mutant, in the presence of sodium azide and 2C4NP-Fuc, did not produce disaccharides, but rather only β-L-Fuc-N_3 (Scheme 24).



Scheme 24 - Chemical rescue of TmD224G using sodium azide.

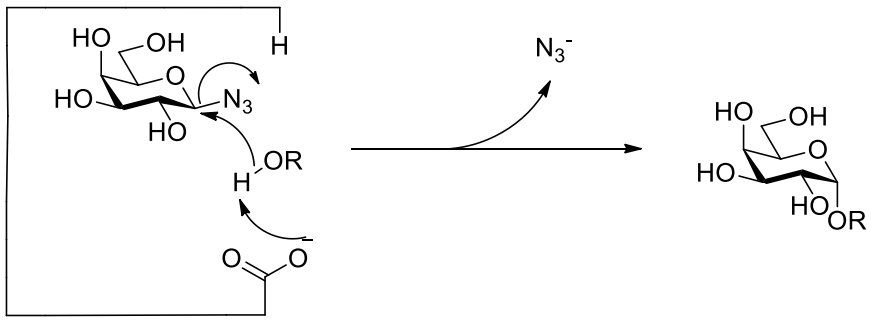
However, in the presence of β-L-Fuc-N_3 and 4NP-β-D-Xyl as the donor and acceptor respectively, the mutant synthesized disaccharides as a mixture of the α1,4 and α1,3 regioisomers, with a fucosylation yields of up to 91 % (Scheme 25).



Scheme 25 - Glycosylation of p-nitrophenyl β-D-xyloside to β-L-Fuc-N_3 by TmD224G.

To demonstrate that the usage of β-glycosyl azides was not limited to α-fucosynthases, Cobucci-Ponzano *et al.* tested β-galactosyl azide (β-Gal-N_3) as a possible donor for some α-galactosynthases.[59] The α-galactosidase from *Thermotoga maritima* (TmGalA, from family GH36) was chosen as a model enzyme. TmGalA follows a retaining

reaction mechanism in which D327 and D387 are the catalytic nucleophile and catalytic acid/base, respectively.[60] Mutants of TmGalA were prepared and the activity of the enzymes was assessed. Whilst the activity of the TmGalA D327A and D327S mutants could not be rescued, the D327G mutant was chemically rescued with sodium azide using 4-nitrophenyl α-D-galactopyranoside (4NP-Gal) by 30-fold compared to the D327G incubated with 4NP-Gal alone. Using the TmGalA D327G mutant, the authors demonstrated galactosylation of a variety of glycosyl acceptors using β-D-Gal-N_3 as the donor, with efficiencies of up to 51 % (Scheme 26).



Scheme 26 - Mechanism of the D327G mutant α-galactosynthases using galactosyl azides as donors. R = Glc-α-pNP, Xyl-α-pNP, Xyl-β-pNP, Man-α-pNP.

It is therefore possible that β-glycosyl azides might act as viable donors for α-glycosynthases generated from a GH101 *N*-acetyl-α-D-galactosaminidase. Glycosyl azides may serve as better donors than the more commonly used glycosyl fluorides for several reasons. As stated before β-glycosyl fluorides have much shorter half-lives with respect to aqueous hydrolysis than α-glycosyl fluorides due to the anomeric effect.[61] Therefore β-glycosyl fluorides are probably too unstable to survive the long reaction times typically required by α-glycosynthases. A substrate with a better balance between stability and reactivity may be a more effective donor for glycosynthase reactions.

3.1.1.4. *p*-Nitrophenyl glycosides

p-Nitrophenyl glycosides are excellent substrates for glycosidases both for analysing their kinetic activity, as well as their utility as substrates for glycosylation reactions. Their utility stems from the leaving group ability of *p*-nitrophenol (pK_a 7.16), and the strong UV chromophore which allows reactions to be easily monitored by HPLC or UV-vis spectroscopy. In a typical reaction, nitrophenyl glycosides are incubated with an enzyme at the desired pH and substrate concentrations for a specified period of time. The reaction is then quenched and the *p*-nitrophenolate ion is spectrophotometrically measured at an alkaline pH, usually 10-11. Alternatively, a continuously monitored spectrophotometric assay[62] can be performed, which is preferable to the former static, fixed time assay due to errors associated with inaccuracies with initial sample mixing times and final sample quenching times. A continuous assay allows for the entire kinetic course of the reaction to be obtained with one small sample.

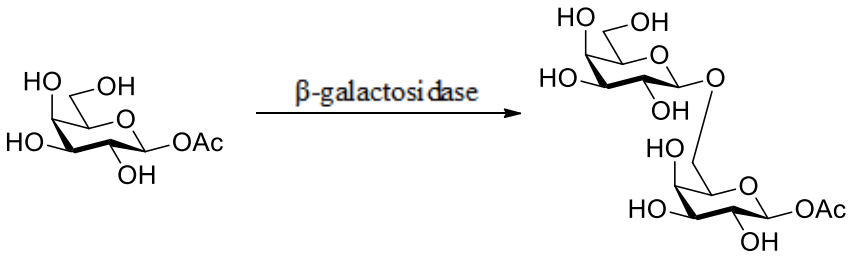
A kinetic assay of glycosidases is typically followed by measuring the increase in absorbance at 318 nm, although for high concentrations slightly longer wavelengths can be used, which reduces the background absorbance and sensitivity ($\Delta\varepsilon$) of the assay.[62] Also, the cleavage of *p*-nitrophenol from a sugar can be monitored due to the differences in the absorptions of *p*-nitrophenol and *p*-nitrophenyl glycosides.[62]

In investigations performed with engBL and engEF, the *p*-nitrophenyl glycosides were used substrates for hydrolytic studies, and as the donors for glycosylation of a set of alkanols.[63,64] In 2011, Hollinger *et al.* published a paper on the synthesis of the *p*-nitro and *p*-aminophenyl glycosides of the mucin *O*-glycan core structures.[65] This paper detailed total syntheses of the expensive core structure units using the inexpensive D-

galactose as the starting material. Work reported in this chapter extends this route to additionally generate glycosyl acetates of these core structures for use in the glycosidase-catalysed glycosylation reactions reported in Chapter 4.

3.1.1.5. Glycosyl acetates

The development of new donors that, in conjunction with glycosidases, results in high glycosylation activity, is essential to increase the yield and efficiency of these processes. Ideally, the donor used should be relatively stable to normal storage conditions, and readily activated when used as a substrate in enzymatic reactions. Of the currently available glycosyl donors, fluoro-, *p*-nitrophenyl, and dinitrophenyl glycosides have been used most extensively. The use of glycosyl acetates as donors has been much less common. The first report on their use was by Zinin *et al.* who, in 2002, demonstrated that a β-galactosidase from *Penicillium* sp. was capable of using 1-*O*-acetyl-β-D-galactopyranoside as a novel donor for glycosylation (Scheme 27). They found that, not only was this β-galactosidase able to catalyse the hydrolysis of 1-*O*-acetyl-β-D-galactopyranoside, but it could also catalyse self-condensation of the donor to give the disaccharide β-D-galactopyranosyl-1,6-1-*O*-acetyl-β-D-galactopyranoside in 70 % yield after 10 mins.



Scheme 27 - Self condensation of 1-*O*-acetyl-β-D-galactopyranoside catalysed by *Penicillium* sp. β-galactosidase.

This observation is not surprising given the low pK_a of acetic acid (4.76), meaning that acetate is a better leaving group than *p*-nitrophenolate. Indeed, it was demonstrated that the catalytic activity of the β-galactosidase from *Penicillium* sp. for the hydrolysis of 1-*O*-acetyl-β-D-galactopyranoside was 100-fold greater than that for *p*-nitrophenyl β-D-galactopyranoside, and 21000-fold greater than that for the equivalent methyl glycoside. Typically, syntheses of the 1-*O*-acetyl donor requires a multi-step synthetic sequence where the acetyl group is installed at the anomeric centre *via* a glycosyl imidate by means of a carboxylate[66] or using AcOH.[67–69]

However since the report by Zinin *et al*[69], there has been no further work published on the usage of glycosyl acetates as donor substrates for glycosidases. This lack of usage may be due to the difficult in synthesizing the required donors as the single anomers required for enzyme reactions, and the low yields in which they can be produced.[68] Therefore, there is a need for a facile route to access deprotected glycosyl acetates. Chapter 5 details a method that allows direct access to glycosyl acetates in water directly from unprotected sugars in high yields, some of which are formed exclusively as single anomers.

3.2. The synthesis of activated *O*-glycan core structures

There are currently 10 core *O*-glycan structures known, each of which differs in their extensions from the *N*-acetyl-D-galactosamine unit at the reducing end (Figure 2).

The two enzymes acquired for use in this thesis were engBL from *Bifidobacterium longum*[63,70,71] and engEF from *Enterococcus faecalis*.[64,72] EngBL has a hydrolytic substrate preference for GalNAc (the Tn antigen) and Galβ1,3GalNAc (the core 1

disaccharide), whilst engEF has a broader substrate specificity, catalysing the hydrolysis of GalNAc, Galβ1,3GalNAc, and GlcNAcβ1,3GalNAc (the core 3 disaccharide).

As explained in Chapter 1, the two main approaches previously employed to enable the enzyme-catalysed formation of the glycosidic bond are thermodynamic controlled reverse hydrolysis and kinetically controlled glycosylation, whereby the latter method is preferred due to higher yielding reactions. One attributing component to the high yields is the use of activated glycosyl donors.

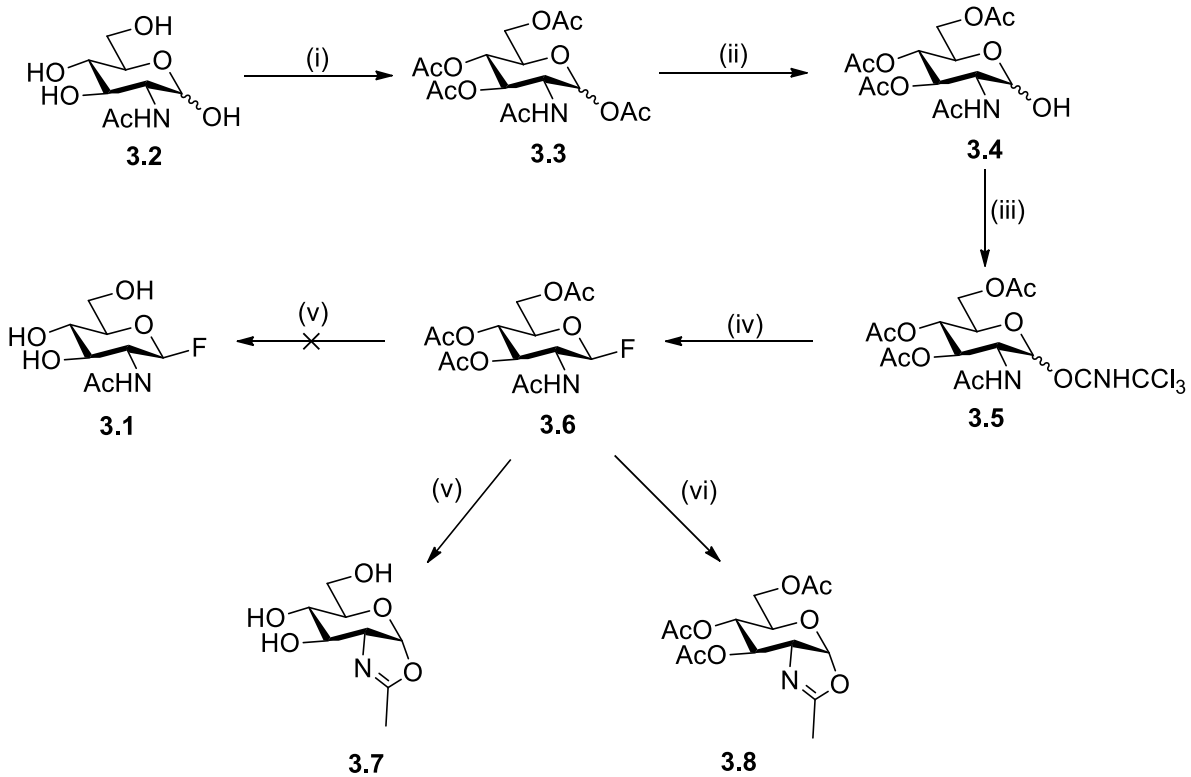
The GH101 endo-α-*N*-acetylgalactosaminidases catalyse the hydrolysis of the GalNAc-α-Ser/Thr linkage *via* the formation of a β-glycosyl-enzyme intermediate. Therefore, a kinetically controlled glycosylation using the GH101 WT enzymes requires the synthesis of activated α-glycosyl donors. Since the GH101 enzymes have substrate specificity for GalNAc, Galβ1,3GalNAc (core 1), and GlcNAcβ1,3GalNAc (core 3), the synthesis of activated α-glycosyl donor derivatives is necessary for their use as substrates in the enzyme-catalysed glycosylation reaction.

The research of this thesis also includes the design of GH101 mutant glycosidases, where the nucleophilic residue of the respective protein had been substituted with a non-nucleophilic residue. Without the ability to form the β-glycosyl-enzyme intermediate, we would expect the mutant glycosidase to be inactive. However, in the presence of an activated β-glycosyl donor, which acts to mimic the β-glycosyl-enzyme intermediate, the enzyme should be able to catalyse the ligation of the donor to an acceptor substrate. For this purpose, activated β-glycosyl donors needed to be

synthesised as substrates for the GH101 glycosynthase enzymes. Suitable activated β-glycosyl donors for an α-glycosidase-derived glycosynthase include β-glycosyl fluorides[73], β-glycosyl azides[52,58], β-*p*-nitrophenyl glycosides[74], and β-glycosyl acetates.

3.2.1. Attempted synthesis of glycosyl fluorides of GalNAc

The synthesis of 2-acetamido-2-deoxy-β-D-glucosaminyl fluoride **3.1** has only been previously reported once in the literature (Scheme 28).[17]



Scheme 28 - Synthesis of β-glycosyl fluorides from 2-acetamido sugars. Conditions: i) Ac_2O, pyridine, rt, 24 h; ii) hydrazine acetate, DMF, rt, 17 h, 72 %; iii) 1,1,1-trichloroacetonitrile, DCM, DBU, 0 °C, 3 h, 93 %; iv) HF (70 % solution in pyridine), DCM, 0 °C, 5 mins, 37 %; v) 1 M NaOMe in MeOH, 1 min; vi) Methylamine or n-propylamine, THF.

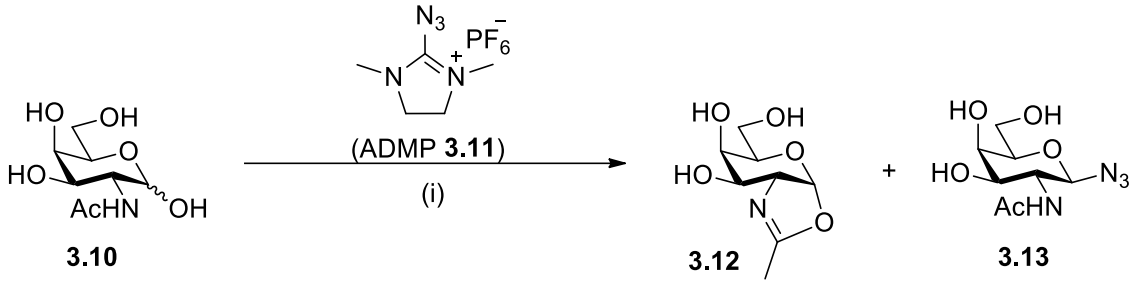
Following the published route[17], GlcNAc **3.2** was acetylated with pyridine and acetic anhydride to give the peracetate **3.3**. Selective anomeric cleavage using hydrazine acetate gave the hemiacetal **3.4**, which was then converted to the imidate **3.5** using

trichloroacetonitrile and DBU. Treatment of imidate **3.5** with HF/pyridine gave the β-glycosyl fluoride **3.6** in 37 % yield. However, Zemplén de-acetylation to yield deprotected fluoride **3.1** using the reported procedure did not occur as described. Instead, it was found that the reaction mixture consisted predominantly of the glycosyl oxazoline **3.7**, indicating that the β-glycosyl fluoride was unstable toward the de-esterification conditions used. The use of either methylamine or n-propylamine for ester cleavage also produced the acetylated glycosyl oxazoline **3.8** as the major product.

After discussion with Prof. Stephen Withers, it was concluded that β-glycosyl fluorides derived from 2-acetamido sugars were probably too unstable to be used as substrates for glycosidase-catalysed synthesis, and therefore route development towards deprotected glycosyl fluorides was abandoned.

3.2.2. Synthesis of 2-acetamido-2-deoxy-β-D-galactopyranosyl azide 3.9

Full details of the development of the synthetic method for the one-pot synthesis of glycosyl azides will be further discussed in Chapter 5. Under optimised conditions, 2-acetamido-2-deoxy-D-galactose **3.10** was reacted with ADMP **3.11** in the presence of triethylamine in D_2O/MeCN (4:1) (Scheme 29). After 3 h, ^{1}H NMR analysis revealed that both oxazoline **3.12** and azide **3.13** were present in the reaction mixture in equal amounts. The oxazoline by-product is formed due to the competing nucleophilic properties of the 2-acetamido functionality which may react with the β-imidazolinium intermediate that is formed by reaction of the β-anomeric hydroxyl group with DMC (see Chapter 5, Figure 46). A simple adjustment of the pH to 2 by dropwise addition of HCl led to the production of the desired azide **3.9** in 77 % yield and the complete disappearance of oxazoline **3.12**.

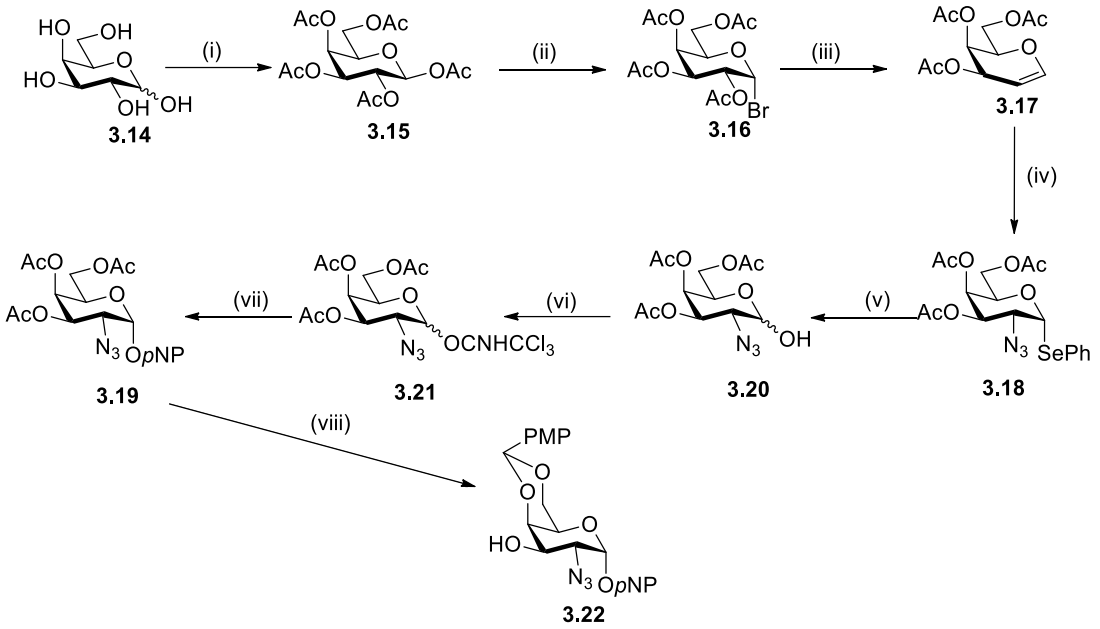


Scheme 29 - (i) a) ADMP, triethylamine, D_2O/MeCN (4:1), 0 °C, 3 h, then HCl to pH 2, 77 % of **3.13**.

Investigations into use of azide **3.13** as a donor for glycosynthase-catalysed glycosylation are reported in Chapter 4. These studies suggested that the glycosynthase mutants of engBL and engEF did not utilize glycosyl azides as substrates. In light of these disappointing results, the syntheses of Galβ1,3GalNAc-β-N_3 (core 1-β-N_3) and the GlcNAcβ1,3GalNAc-β-N_3 (core 3-β-N_3) glycosyl donors were not undertaken.

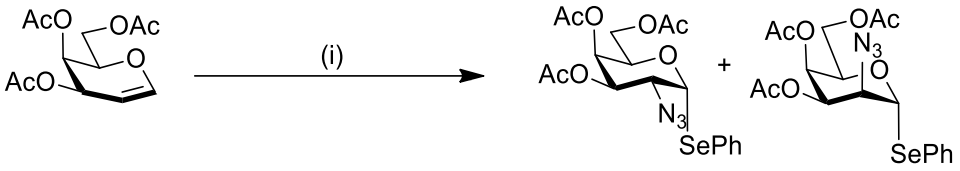
3.2.3. The synthesis of *p*-nitrophenyl β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-α-D-galactopyranoside 3.13 (core 1 α-pNP) and *p*-nitrophenyl 2-acetamido-2-deoxy-β-D-glucopyranosyl-(1→3)-2-acetamido-2-deoxy-α-D-galactopyranoside 3.14 (core 3 α-pNP) disaccharides

Scheme 30 illustrates the route used to access the common acceptor **3.22** used for the synthesis of the core *O*-glycan disaccharides.



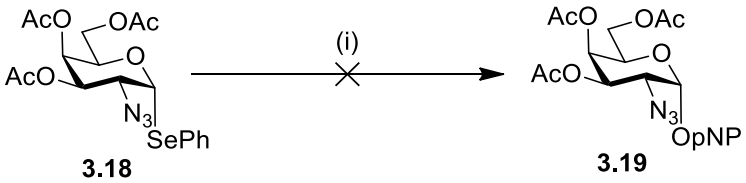
Scheme 30 - Synthetic route towards core 1-pNP. Conditions (i) Ac_2O, NaOAc, 120 °C, 1 h, 37 %; (ii) 33 % HBr in AcOH, DCM, 3 h, 100 %; (iii) Zn dust, 2,6-lutidine, EtOAc, 60 °C, 2 h, 79 %; (iv) BAIB, NaN_3, $(PhSe)_2$, DCM, 17 h, 40 %; (v) NIS, THF:H_2O (1:1), 5 h, 99 %; (vi) Cl_3CCN, DBU, DCM, 0 °C, 3 h, 98 %; (vii) pNP-OH, TMSOTf, DCM, 4Å MS, 3.5 h, 76 %; (viii) 1) NaOMe, MeOH; 2) 4-methoxybenzaldehyde dimethyl acetal, *p*-TSA, DMF, MeCN, 24 h, 37 %.

Firstly, D-galactose **3.14** was acetylated with sodium acetate and acetic anhydride to give exclusively the β-pentaacetate **3.15**, which was then converted to the α-bromide **3.16** using hydrogen bromide in AcOH in quantitative yield. Reductive elimination using zinc dust then afforded the triacetyl galactal **3.17** in 79 % yield. Azidophenylselenation was performed with sodium azide and diphenyldiselenide in the presence of bis(acetoxy)iodobenzene (BAIB).[75] As expected, a mixture of stereoisomers was observed, all of which had similar polarities by TLC and were inseparable by column chromatography (Scheme 31).



Scheme 31 - Stereoisomers formed during the azidophenylselenation of galactal. Conditions: NaN_3, BAIB, $(PhSe)_2$, DCM.

However the desired product could easily be crystallised from the crude mixture to give selenoglycoside **3.18** in 40 % yield. An attempt was made to directly convert selenoglycoside **3.18** to the *p*-nitrophenyl glycoside **3.19** using NIS and *p*-nitrophenol, unfortunately this was unsuccessful, only yielding the hydrolysed starting material (Scheme 32).



Scheme 32 - Attempted glycosylation of *p*-nitrophenol to selenoglycoside **3.18**. Conditions: *p*-nitrophenol, NIS, DCM.

An alternative route was devised wherein the selenoglycoside was hydrolysed by treatment with NIS and water to give the hemiacetal **3.20**, which was converted to the imidate **3.21** using Schmidt's method.[76] Activation of the imidate with TMSOTf in the presence of *p*-nitrophenol as the acceptor afforded the desired *p*-nitrophenyl glycoside **3.19**, as the α-anomer exclusively. Zemplén deacetylation of triacetate **3.19** followed by treatment with anisaldehyde dimethyl acetal and *p*-toluene sulfonic acid gave the *p*-methoxybenzylidene **3.22**, which was used as the common acceptor unit for the synthesis of the core 1 and core 3 disaccharides.

The non-reducing sugar of the core 1 disaccharide is galactose which is linked to the *N*-acetyl-D-galactosamine *via* a β1,3 linkage. A galactose donor suitable for glycosylation of acceptor **3.22** was synthesised by treating galactose **3.14** with benzoyl chloride in pyridine to obtain the perbenzoylated galactoside **3.23** (Scheme 33). It is well established that glycosyl donors with an ester functionality at the 2-position give exclusively the 1,2-*trans* glycoside. Normally, acetyl protecting groups are used at the 2-position to facilitate neighbouring group participation during glycosylation. However in this case, the corresponding peracetylated donor was found to favour orthoester formation during the glycosylation reaction.[65] Using benzoyl esters mitigated this issue (Figure 29).

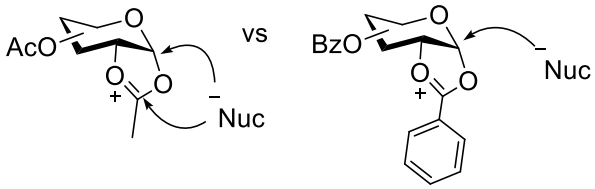
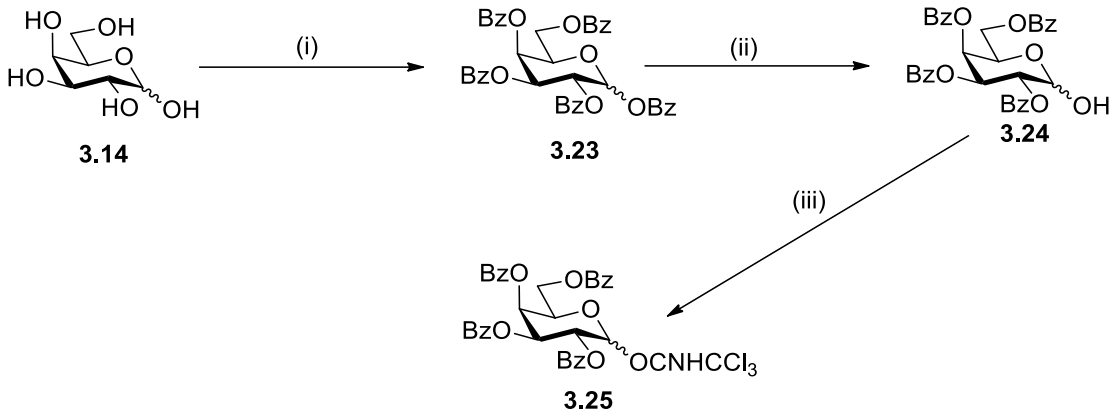


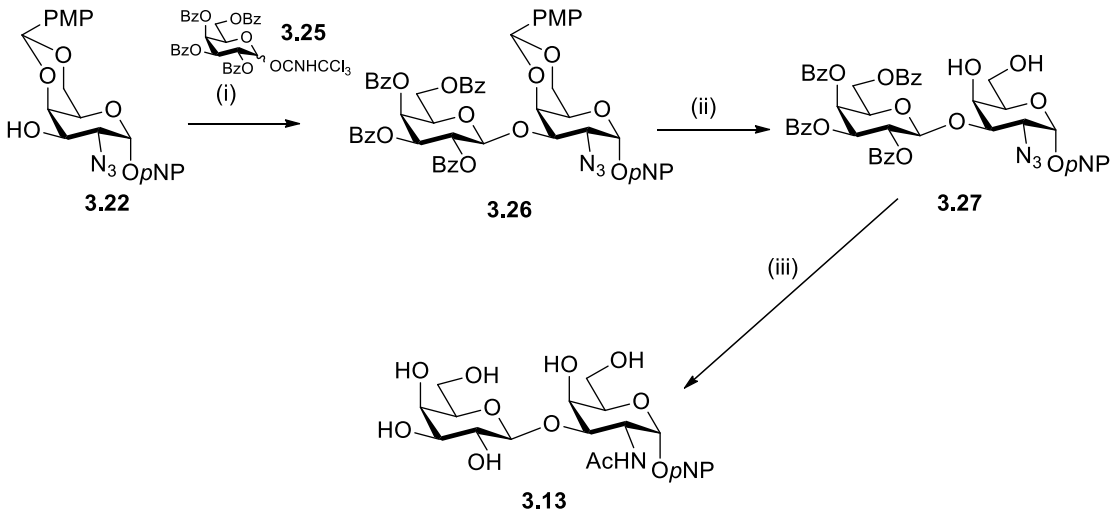
Figure 29 - Neighbouring Group Participation a C-2 acetyl ester compared with a benzoyl ester. Nucleophilic attack to form the orthoester product is prevented by the decreased electrophilicity of the acetoxonium carbon due to resonance stabilisation by the benzyl group.

Benzoyl glycoside **3.23** was then regioselectively deprotected with hydrazine acetate to give galactopyranose **3.24**, which was converted to the galactosyl imidate **3.25** (Scheme 33).



Scheme 33 - Synthesis of galactosyl donor **3.25.** Conditions: i) D-galactose, benzoyl chloride, pyridine, 0 ºC → rt, 18 h, 79 %; ii) hydrazine acetate, DMF, rt, 5.5 h, 95 %; iii) 1,1,1-trichloroacetonitrile, DBU, DCM, 0 ºC, 5 h, 81 %.

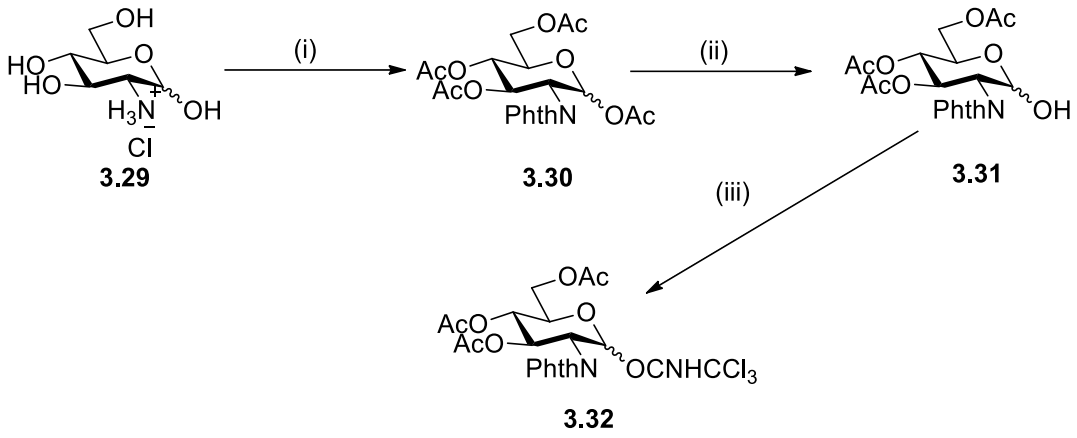
Acceptor **3.22** was then coupled with the trichloroacetimidate donor **3.25** under Lewis acid catalysis to give the desired disaccharide **3.26** in 45 % yield (Scheme 34). Hydrolysis of the 4,6-benzylidene with 80 % acetic acid then gave the diol **3.27**. The azide was converted to the required acetamide in one step using thioacetic acid. Subsequent Zemplén deacetylation of the crude product from this reaction then gave the desired core 1-pNP glycoside **3.13**.



Scheme 34 - Synthesis of core 1-pNP. Conditions: i) **3.25**, TMSOTf, DCM, 4Å MS, -30 ºC, 4 h, 45 %; ii) 80 % aq. AcOH, DCM, rt, 19 h, 85 %; (iii) 1) AcSH, pyridine, rt, 36 h; 2) NaOMe, MeOH, rt, O/N, 9 %.

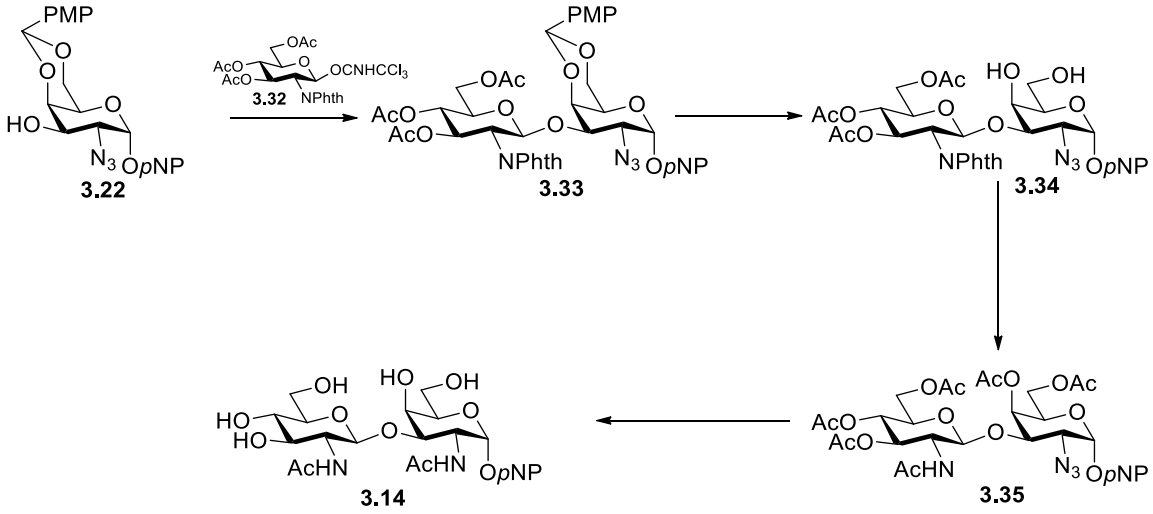
The synthesis of the core 3 disaccharide was also achieved using the common core monosaccharide acceptor **3.22**. However, an *N*-acetyl-D-glucosamine donor also needed to be synthesised. Hollinger *et al*. had previously reported glycosylation of acceptor **3.22**, using a GlcNAc thioglycoside. However, we decided to use a trichloroacetimidate donor so as to avoid the use of malodorous thiols.

The required donor was easily synthesised in three steps as shown in Scheme 33. Firstly, the amine of glucosamine hydrochloride **3.29** was protected with a phthalimide functionality using phthalic anhydride, which was immediately followed by esterification of the remaining hydroxyl groups by treatment with acetic anhydride to give fully protected glucosaminide **3.30** (Scheme 35). Glycosyl acetate **3.30** was regioselectively deprotected using hydrazine acetate in DMF to yield the free hemiacetal **3.31**, which was then converted to the corresponding imidate donor **3.32**.



Scheme 35 - Synthesis of glycosyl donor **3.32**. Conditions: i) 1) Phthalic anhydride, NaOH, methanol, water, acetone, $NaHCO_3$, 25 h, rt; 2) Acetic anhydride, pyridine, rt, O/N, 75 %; ii) hydrazine acetate, DMF, rt, 23 h, 69 %; iii) 1,1,1-trichloroacetonitrile, DBU, DCM, 0 ºC, 6 h, 93 %.

Imidate **3.32** (2.3 equiv) was activated with TMSOTf at -30 ºC in the presence of acceptor **3.22** (Scheme 36). After 7 h, TLC indicated that the acceptor had been completely consumed and that a new product had formed with a polarity very similar to the donor, which would have made isolation of disaccharide **3.33** very difficult. The addition of water to the reaction mixture then hydrolysed any remaining imidate **3.32** and give disaccharide **3.33** in 68 % yield. The *p*-methoxybenzylidene was then hydrolysed using 80 % aqueous AcOH to give the diol **3.34**. The phthalimide protecting group was subsequently converted to an acetamide **3.35** in three steps using sodium methoxide, then hydrazine acetate, followed by subsequent acetylation using acetic anhydride. The azido functionality of **3.35** was then converted in one step to an acetamide using thioacetic acid, and this was followed by Zemplén deacetylation of the crude intermediate to give the desired core 3-pNP disaccharide **3.14**.

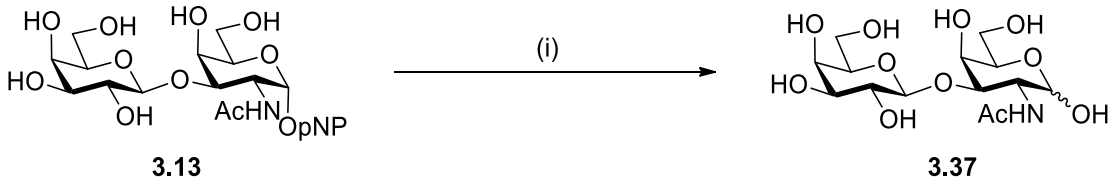


Scheme 36 - Synthetic route towards core 3-pNP. Reactions and conditions: (i) **3.32**, TMSOTf, DCM, 4 Å MS, -30 °C, 7 h, 72 %; (ii) 80 % AcOH, DCM, 44 h, 68 %; (iii) 1) NaOMe, MeOH, O/N; 2) Hydrazine acetate, EtOH, 70 °C, 18 h; 3) Ac_2O, pyridine, DMAP, O/N, 80 %; (iv) 1) AcSH, pyridine; 2) NaOMe, MeOH, 11 %.

3.2.4. The synthesis of 1-*O*-acetyl core structures

To access glycosyl acetate donors, it was necessary to generate the glycosyl hemiacetals from the core disaccharide structures. The hydrolysis of the Galβ1,3GalNAc-α-pNP (core 1-pNP) and the GlcNAcβ1,3GalNAc-α-pNP (core 3-pNP) glycosides were easily performed using the well reported hydrolytic capabilities of engEF from *Enterococcus faecalis*.[64]

To this end, core 1-pNP **3.13** was dissolved in the appropriate buffer and treated with engEF (Scheme 37). Hydrolysis was monitored by RP-HPLC, and was found to be complete after 1 h. Purification by RP-HPLC gave disaccharide **3.37** in 99 % yield.



3.13 **3.37**

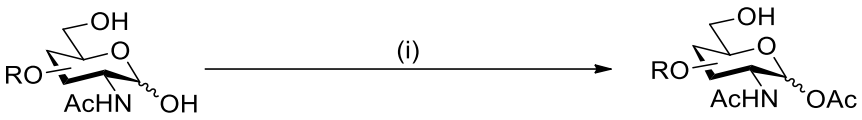
Scheme 37 - Hydrolysis of core 1-pNP. Conditions: i) engEF from *Enterococcus faecalis*, 25 mM NaOAc, pH 6.0, 37 ºC, 1 h, 99 %.

In a similar manner, the core 3-pNP glycoside was also hydrolysed using engEF to give the disaccharide hemiacetal **3.38** in 45 % yield (Scheme 38).

3.14 **3.38**

Scheme 38 - Hydrolysis of core 3-pNP. Conditions: i) engEF from *Enterococcus faecalis*, 25 mM NaOAc, pH 6.0, 37 ºC, 1 h, 45 %.

Two of the core *O*-glycan structures were then converted to their 1-*O*-acetates for use in later enzymatic investigations (Table 7). Firstly, *N*-acetyl-D-galactosamine **3.9** was converted to the 1-*O*-acetate **3.36** using 2-chloro-1,3-dimethylimidazolidinium chloride (DMC) and TEA in 86 % yield as an anomeric mixture (α/β, 1:1.1, Table 7, Entry 1).



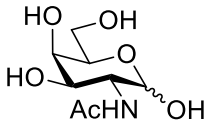
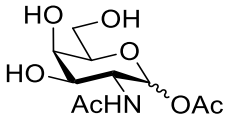
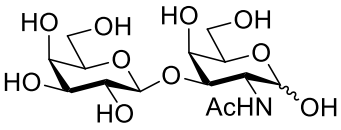
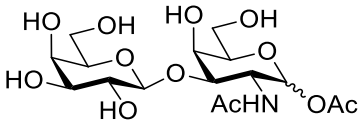
Entry	Sugar	Product	Yield (α/β)
1	 **3.9**	 **3.36**	86 % (1:1.1)
2	 **3.37**	 **3.39**	87 % (1.3:1)

Table 7: Conversion of *O*-glycan-associated structures to their 1-*O*-acetates. Conditions: (i) sugar, TEA, AcSH, H_2O, 0 ºC, then DMC, 1 h.

The formation of a mixture of anomers can be rationalized by the consideration of the putative mechanism of the reaction (Figure 30). Firstly, thioacetate, generated by the deprotonation of thioacetic acid by triethylamine, reacts with DMC to generate a reactive thioester intermediate. This thioester is then attacked by the anomeric hydroxyl group of *N*-acetyl-D-galactosamine **3.9**. Since **3.9** exists as a mixture of anomers in solution, then both anomers can react with the intermediate leading to a mixture of acetylated anomers.

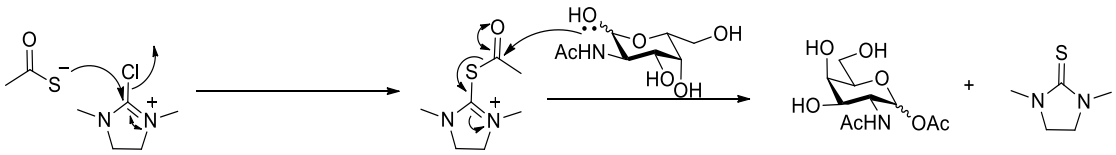


Figure 30 - Mechanism of glycosyl acetate formation using DMC and thioacetic acid.

However, in order to be used as substrates for enzyme catalysed glycosylations, it was preferable that pure anomers were obtained. The presence of both anomers is not

evidenced by TLC because of very small differences in their R_f values, meaning an alternative to normal phase chromatography was necessary. It was found that both anomers were easily separable using RP-HPLC to obtain the pure α and β anomers (Figure 31).

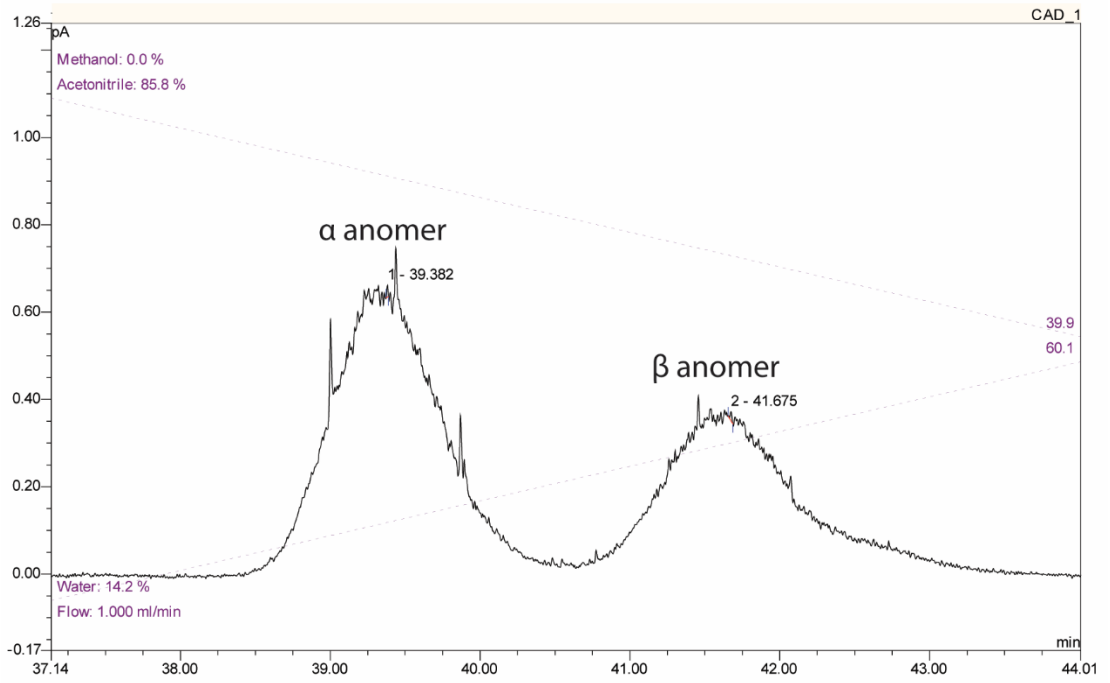


Figure 31 - HPLC trace of the separation of GalNAc-1-*O*-acetate **3.36** anomers by RP-HPLC. Detector: CAD.

The core 1 disaccharide was also converted to the corresponding 1-*O*-acetate using the same method, which again produced the acetates as a mixture of anomers (Table 7, Entry 2). These were again easily separated by RP-HPLC (Figure 32) to give the desired product as the pure single anomers in 87 % overall yield.

The revelation that the GH101 mutant ENGases, which are discussed in Chapter 4, were in fact inactive towards the core 1 disaccharide glycosyl acetates meant that a decision was made not to also synthesise the corresponding core 3 glycosyl acetate.

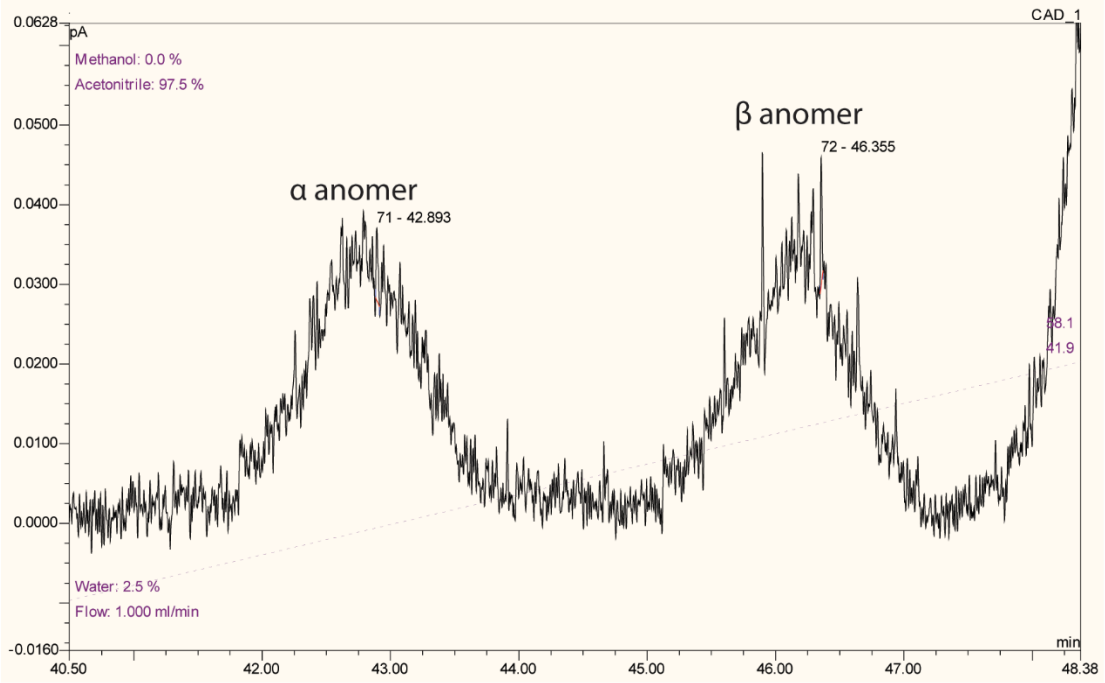


Figure 32 - HPLC trace for the separation and isolation of core 1-OAc anomers. Detector: CAD.

3.3. Conclusions

Herein, activated substrates for the GH101 *N*-acetyl-α-D-galactosaminidases were synthesised. GalNAc-α-OAc, Gal1,3GalNAc-α-pNP (core 1-pNP), and GlcNAcβ1,3GalNAcαpNP (core 3-pNP) were synthesised as substrates for the WT engBL and engEF enzymes whilst GalNAc-β-OAc, GalNAc-β-N_3, and Galβ1,3GalNAc-β-OAc were synthesised for the mutant engBL and engEF enzymes. An attempt was made to synthesise activated β-glycosyl fluoride derivatives of the core structures using a published method however this was unsuccessful due to the instability of the β-fluoride towards de-esterification conditions. Also, due to disappointing results of their usage as enzymatic substrates (Chapter 5), the β-glycosyl azide derivatives of the core disaccharide structures were not synthesised.

The compounds synthesised in this Chapter will be used as substrates for assaying the kinetic activity of these enzymes as well as being used for enzyme-catalysed glycosylation of acceptors in Chapter 4.

3.4. References

(1) Villalonga, M. L., Díez, P., Sánchez, A., Gamella, M., Pingarrón, J. M., Villalonga, R. *Chem. Rev.* **2014**, *114*, 4868-4917.

(2) Wang, Z., Chinoy, Z. S., Ambre, S. G., Peng, W., McBride, R., de Vries, R. P., Glushka, J., Paulson, J. C., Boons, G. J. *Science*. **2013**, *341*, 379-783.

(3) Yamamoto, K.; Kadowaki, S.; Watanabe, J.; Kumagai, H. *Biochem. Biophys. Res. Commun.* **1994**, *203* (1), 244–252.

(4) Fialová, P.; Weignerová, L.; Rauvolfová, J.; Přikrylová, V.; Pišvejcová, A.; Ettrich, R.; Kuzma, M.; Sedmera, P.; Křen, V. *Tetrahedron* **2004**, *60* (3), 693–701.

(5) Wang, L. X.; Huang, W. *Curr. Opin. Chem. Biol.* **2009**, *13* (5-6), 592–600.

(6) Koshland, D. E. *Biol. Rev.* **1953**, *28* (4), 416–436.

(7) Schmaltz, R. M.; Hanson, S. R.; Wong, C. H. *Chem. Rev.* **2011**, *111* (7), 4259–4307.

(8) Perugino, G.; Trincone, A.; Rossi, M.; Moracci, M. *Trends Biotechnol.* **2004**, *22* (1), 31–37.

(9) Cobucci-Ponzano, B.; Strazzulli, A.; Rossi, M.; Moracci, M. *Adv. Synth. Catal.* **2011**, *353* (13), 2284–2300.

(10) Williams, S. J.; Withers, S. G. *Carbohydr. Res.* **2000**, *327* (1-2), 27–46.

(11) Micheel, F.; Klemer, A. *Adv. Carbohydr. Chem.* **1962**, *16*, 85–103.

(12) Hayashi, M.; Hashimoto, S.; Noyori, R. *Chem. Lett.* **1984**, *10*, 1747–1750.

(13) Igarashi, K.; Honma, T.; Irisawa, J. *Carbohydr. Res.* **1970**, *13* (1), 49–55.

(14) Yokoyama, M. *Carbohydr. Res.* **2000**, *327* (1-2), 5–14.

(15) Furuya, T.; Kuttruff, C. A; Ritter, T. *Curr. Opin. Drug Disc.* **2008**, *11* (6), 803–819.

(16) Konstantinidis, A; Sinnott, M. L. *Biochem. J.* **1991**, *279* (2), 587–593.

(17) Ballardie, F. W.; Capon, B.; Dearie, W. M.; Foster, R. L. *Carbohydr. Res.* **1976**, *49* (1576), 79–92.

(18) Copa-Patiño, J. L.; Zhang, Y.; Padmaperuma, B.; Marsden, I.; Broda, P.; Sinnott, M. L. *Biochem. J.* **1993**, *293* (2), 591–594.

(19) Knowles, J. K. C.; Lentovaara, P.; Murray, M.; Sinnott, M. L. *J. Chem. Soc. Chem. Commun.* **1988**, *21*, 1401.

(20) Hehre, E. J.; Brewer, C. F.; Genghof, D. S. *J. Biol. Chem.* **1979**, *254* (13), 5942–5950.

(21) Kasumi, T.; Tsumuraya, Y.; Brewer, C. F.; Kersters-Hilderson, H.; Claeyssens, M.; Hehre, E. J. *Biochemistry* **1987**, *26* (11), 3010–3016.

(22) Kasumi, T.; Brewer, C. F.; Reese, E. T.; Hehre, E. J. *Carbohydr. Res.* **1986**, *146* (1), 39–42.

(23) Kitahata, S.; Brewer, C. F.; Genghof, D. S.; Sawai, T.; Hehre, E. J. *J. Biol. Chem.* **1981**, *256* (12), 6017–6026.

(24) Hehre, E. J.; Sawai, T.; Brewer, C. F.; Nakano, M.; Kanda, T. *Biochemistry* **1982**, *21* (13), 3090–3097.

(25) Withers, S. G.; Aebersold, R. *Protein Sci.* **1995**, *4* (3), 361–372.

(26) Vocadlo, D. J.; Davies, G. J.; Laine, R.; Withers, S. G. *Nature* **2001**, *412* (6849), 835–838.

(27) Braun, C.; Brayer, G. D.; Withers, S. G. *J. Biol. Chem.* **1995**, *270* (45), 26778–26781.

(28) McCarter, J. D.; Withers, S. G. *J. Am. Chem. Soc.* **1996**, *118* (1), 241–242.

(29) Matsui, H.; Blanchard, J. S.; Brewer, C. F.; Hehre, E. J. *J. Biol. Chem.* **1989**, *264* (15), 8714–8716.

(30) Tanaka, Y.; Tao, W.; Blanchard, J. S.; Hehre, E. J. *J. Biol. Chem.* **1994**, *269* (51), 32306–32312.

(31) Barnett, J. E.; Jarvis, W. T.; Munday, K. A. *Biochem. J.* **1967**, *103* (3), 699–704.

(32) Hehre, E. J.; Genghof, D. S.; Okada, G. *Arch. Biochem. Biophys.* **1971**, *142* (1), 382–393.

(33) Prade, H.; Mackenzie, L. F.; Withers, S. G. *Carbohydr. Res.* **1997**, *305* (3-4), 371–381.

(34) Viladot, J.-L.; Moreau, V.; Planas, A.; Driguez, H. *J. Chem. Soc. Perkin Trans. 1* **1997**, *16*, 2383–2388.

(35) Viladot, J.-L.; Stone, B.; Driguez, H.; Planas, A. *Carbohydr. Res.* **1998**, *311* (1-2), 95–99.

(36) Hrmova, M.; Fincher, G. B.; Viladot, J.-L.; Planas, A.; Driguez, H. *J. Chem. Soc. Perkin Trans. 1* **1998**, *21*, 3571–3576.

(37) Karthaus, O.; Shoda, S.-I.; Takano, H.; Obata, K.; Kobayashi, S. *J. Chem. Soc. Perkin Trans. 1* **1994**, *13*, 1851.

(38) Svensson, S. C. T.; Thiem, J. *Carbohydr. Res.* **1990**, *200*, 391–402.

(39) Lougheed, B.; Ly, H. D.; Wakarchuk, W. W.; Withers, S. G. *J. Biol. Chem.* **1999**, *274* (53), 37717–37722.

(40) Mackenzie, L. F.; Wang, Q.; Warren, R. A. J.; Withers, S. G. *J. Am. Chem. Soc.* **1998**, *120* (22), 5583–5584.

(41) Capon, B.; Thomson, J. W. *Bioorg. Chem.* **1979**, *8* (2), 147–173.

(42) Ganguli, A. R. S.; Coward, J. K. *Tetrahedron: Asymmetry* **2005**, *16* (2), 411–424.

(43) Lee, S. S.; Yu, S.; Withers, S. G. *Biochemistry* **2003**, *42* (44), 13081–13090.

(44) Zechel, D. L.; Boraston, A. B.; Gloster, T.; Boraston, C. M.; Macdonald, J. M.; Tilbrook, D. M. G.; Stick, R. V.; Davies, G. J. *J. Am. Chem. Soc.* **2003**, *125* (47), 14313–14323.

(45) Vallmitjana, M.; Ferrer-Navarro, M.; Planell, R.; Abel, M.; Ausín, C.; Querol, E.; Planas, A.; Pérez-Pons, J.-A. *Biochemistry* **2001**, *40* (20), 5975–5982.

(46) Koeners, H. J.; de Kok, A. J.; Romers, C.; van Boom, J. H. *Recl. Trav. Chim. Pay. B* **2010**, *99* (11), 355–362.

(47) Berven, L. A.; Dolphin, D. H.; Withers, S. G. *J. Am. Chem. Soc.* **1988**, *110* (14), 4864–4866.

(48) Berven, L. A.; Dolphin, D.; Withers, S. G. *Can. J. Chem.* **1990**, *68* (10), 1859–1866.

(49) Bolam, D. N.; Charnwood, S. J.; Gilbert, H. J.; Hughes, N. A. *Carbohydr. Res.* **1998**, *312* (1-2), 85–89.

(50) Sharma, S. *Tetrahedron Lett.* **1995**, *36* (31), 5627–5630.

(51) Chen, H. M.; Withers, S. G. *ChemBioChem* **2007**, *8* (7), 719–722.

(52) Fialová, P.; Carmona, A. T.; Robina, I.; Ettrich, R.; Sedmera, P.; Přikrylová, V.; Petrásková-Hušáková, L.; Křen, V. *Tetrahedron Lett.* **2005**, *46* (50), 8715–8718.

(53) Cobucci-Ponzano, B.; Trincone, A.; Giordano, A.; Rossi, M.; Moracci, M. *J. Biol. Chem.* **2003**, *278* (17), 14622–14631.

(54) Cobucci-Ponzano, B.; Conte, F.; Mazzone, M.; Bedini, E.; Corsaro, M. M.; Rossi, M.; Moracci, M. *Biocatal. Biotransform.* **2008**, *26* (1-2), 18–24.

(55) Ly, H. D.; Withers, S. G. *Annu. Rev. Biochem.* **1999**, *68*, 487–522.

(56) Sulzenbacher, G.; Bignon, C.; Nishimura, T.; Tarling, C. A.; Withers, S. G.; Henrissat, B.; Bourne, Y. *J. Biol. Chem.* **2004**, *279* (13), 13119–13128.

(57) Tarling, C. A.; He, S.; Sulzenbacher, G.; Bignon, C.; Bourne, Y.; Henrissat, B.; Withers, S. G. *J. Biol. Chem.* **2003**, *278* (48), 47394–47399.

(58) Cobucci-Ponzano, B.; Conte, F.; Bedini, E.; Corsaro, M. M.; Parrilli, M.; Sulzenbacher, G.; Lipski, A.; Dal Piaz, F.; Lepore, L.; Rossi, M.; Moracci, M. *Chem. Biol.* **2009**, *16* (10), 1097–1108.

(59) Cobucci-Ponzano, B.; Zorzetti, C.; Strazzulli, A.; Carillo, S.; Bedini, E.; Corsaro, M. M.; Comfort, D. A.; Kelly, R. M.; Rossi, M.; Moracci, M. *Glycobiology* **2011**, *21* (4), 448–456.

(60) Comfort, D. A.; Bobrov, K. S.; Ivanen, D. R.; Shabalin, K. A.; Harris, J. M.; Kulminskaya, A. A.; Brumer, H.; Kelly, R. M. *Biochemistry* **2007**, *46* (11), 3319–3330.

(61) Juaristi, E.; Cuevas, G. *Tetrahedron* **1992**, *48* (24), 5019–5087.

(62) Ford, J. R.; Nunley, J. A.; Li, Y.-T.; Chambers, R. P.; Cohen, W. *Anal. Biochem.* **1973**, *54* (1), 120–128.

(63) Fujita, K.; Oura, F.; Nagamine, N.; Katayama, T.; Hiratake, J.; Sakata, K.; Kumagai, H.; Yamamoto, K. *J. Biol. Chem.* **2005**, *280* (45), 37415–37422.

(64) Goda, H. M.; Ushigusa, K.; Ito, H.; Okino, N.; Narimatsu, H.; Ito, M. *Biochem. Biophys. Res. Commun.* **2008**, *375* (4), 541–546.

(65) Hollinger, M.; Abraha, F.; Oscarson, S. *Carbohydr. Res.* **2011**, *346* (12), 1454–1466.

(66) Ziegler, T. *Liebigs Ann. der Chemie* **1990**, *1990* (11), 1125–1131.

(67) Wang, R.; Steensma, D. H.; Takaoka, Y.; Yun, J. W.; Kajimoto, T.; Wong, C.-H. *Bioorg. Med. Chem.* **1997**, *5* (4), 661–672.

(68) Veeneman, G. H.; van Leeuwen, S. H.; van Boom, J. H. *Tetrahedron Lett.* **1990**, *31* (9), 1331–1334.

(69) Zinin, A. I.; Eneyskaya, E. V; Shabalin, K. A; Kulminskaya, A. A; Shishlyannikov, S. M.; Neustroev, K. N. *Carbohydr. Res.* **2002**, *337* (7), 635–642.

(70) Suzuki, R.; Katayama, T.; Kitaoka, M.; Kumagai, H.; Wakagi, T.; Shoun, H.; Ashida, H.; Yamamoto, K.; Fushinobu, S. *J. Biochem.* **2009**, *146* (3), 389–398.

(71) Ashida, H.; Ozawa, H.; Fujita, K.; Suzuki, S.; Yamamoto, K. *Glycoconj. J.* **2010**, *27* (1), 125–132.

(72) Koutsioulis, D.; Landry, D.; Guthrie, E. P. *Glycobiology* **2008**, *18* (10), 799–805.

(73) Yamamoto, K.; Davis, B. G. *Angew. Chem. Int. Ed.* **2012**, *51* (30), 7449–7453.

(74) Jahn, M.; Withers, S. G. *Biocatal. Biotransfor.* **2003**, *21* (4-5), 159–166.

(75) Czernecki, S.; Randriamandimby, D. *Tetrahedron Lett.* **1993**, *34* (49), 7915–7916.

(76) Schmidt, R. R.; Michel, J. *Tetrahedron Lett.* **1984**, *25* (8), 821–824.

Chapter 4 – Investigations into the glycosylation activity of GH101 endo-α-*N*-acetylgalactosaminidases

4.1. Introduction

4.1.1. Carbohydrate processing enzymes

The use of enzymes to synthesise glycosidic linkages is an area of research that has received considerable attention in recent years. The ability to stereoselectively, and sometimes also regioselectively, catalyse reactions in water without the need of a protection strategy are just a few of the advantages associated with this method of catalysis. Over recent years, both glycosyltransferases and glycosidases have been used extensively for the construction of glycosidic bonds.[1–3]

Glycosyltransferases are often used for the assembly of oligosaccharides and glycopeptides, and additionally have also been applied to the synthesis of a small number of glycoproteins.[4,5] In general, glycosyltransferases act in a stepwise manner and are highly specific in their substrate requirements.[6,7] However, the requirement for complex sugar nucleotide donors and the lack of commercially available enzymes makes the use of glycosyltransferases unattractive. Moreover, the generation of nucleoside diphosphates as side products, which can inhibit the enzyme, is also problematic.[8]

Glycosyl hydrolases, on the other hand, are plentiful in nature and readily hydrolyse glycosidic bonds of numerous oligosaccharides. To date, well over 260,000 different glycosyl hydrolases have been identified, and, based on sequence analysis, grouped into 135 families.

The hydrolysis of the glycosidic bond by glycosidases can give two possible stereochemical outcomes; either retention or inversion of configuration. A detailed analysis of the active site mechanism of retaining and inverting glycosidases is detailed in Chapter 1.

4.1.2. Mechanism of family GH101 endo-α-*N*-acetylgalactosaminidases

In 2005, Fujita *et al.* discovered an enzyme that catalysed the liberation of Galβ1,3GalNAc (core 1) from mucin-type glycoproteins.[9] The enzyme was purified from *Bifidobacterium longum* and was the first enzyme to be allocated into family GH101.

Since that time, several other enzymes have been discovered and classed as family GH101. The GH101 enzymes catalyse the hydrolysis of their respective substrates by means of a retaining mechanism. As described in Chapter 1, retaining glycosidases use a double-displacement mechanism involving the formation of a glycosyl-enzyme intermediate. This intermediate is then hydrolysed by general base-catalysed attack of water upon the anomeric centre, forming the hemiacetal product with retention of configuration, and returning the enzyme to its original protonation state (Figure 33).

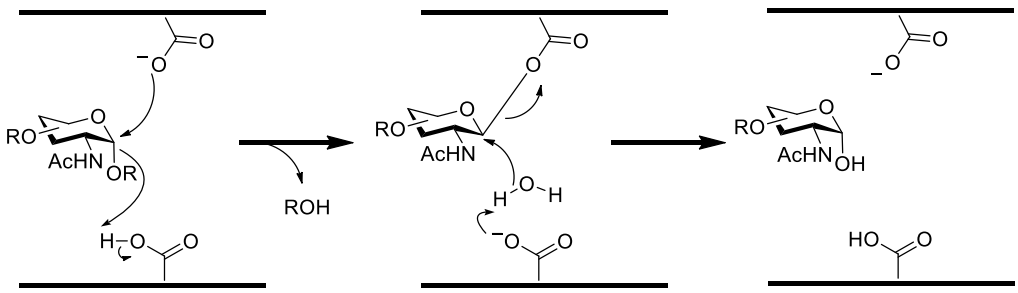


Figure 33 – Hydrolysis mechanism of family GH101 endo-α-*N*-acetylgalactosaminidases.

4.1.3. Glycosynthases

In 1998, a number of research groups investigated the mutagenesis of catalytic active site residues in glycosidases and their applications in oligosaccharide synthesis.[1,10] These studies on β-glycosidases had revealed that site-directed mutagenesis of the catalytic nucleophile to a non-nucleophilic amino acid (such as alanine) removed hydrolytic activity. The importance of the mutation can be appreciated if one considers the mechanism of the retaining β-glycosidase (Figure 34). As mentioned in Chapter 1, retaining glycosidases operate by a double-displacement mechanism involving the formation of a glycosyl-enzyme intermediate (Figure 34a, Step 1).

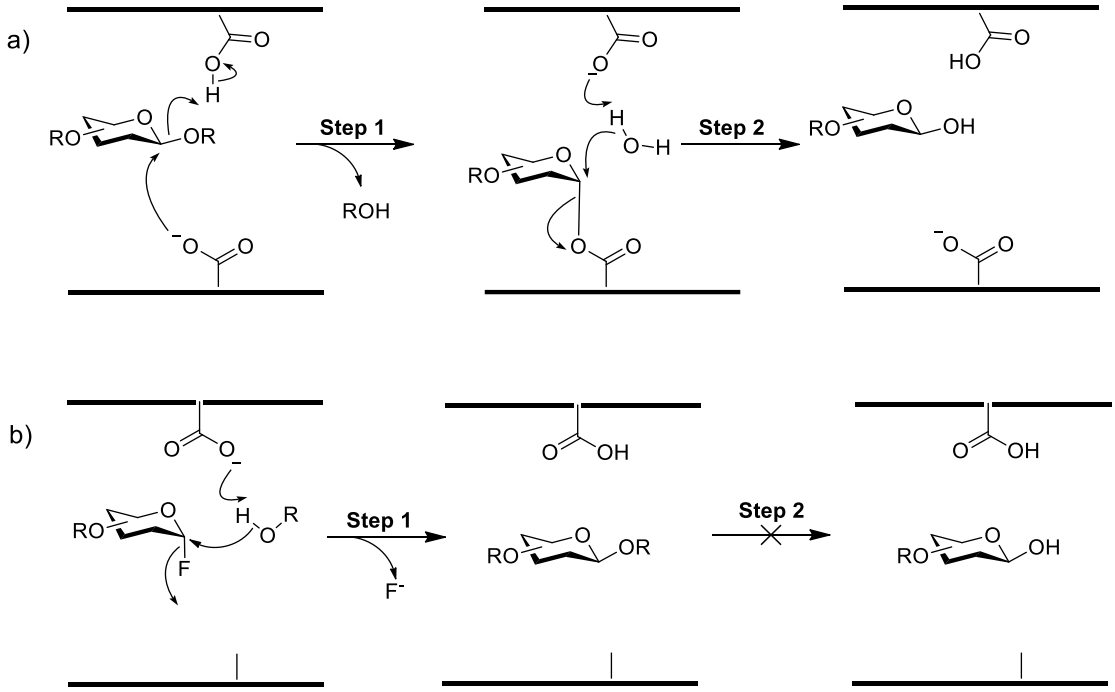


Figure 34 - Mechanisms of a) retaining β-glycosidase and; b) β-glycosynthase catalysis

General base-catalysed attack of an appropriate acceptor (such as water, Figure 34a, Step 2) then forms the product with retention of configuration. Removal of the catalytic nucleophile disrupted the natural mechanism of the enzyme (Figure 34b). Without the ability to form the glycosyl-enzyme intermediate, the glycosidase could not catalyse

the hydrolysis of its natural substrates. However, with an intact active site, it could still catalyse the ligation of an activated α-glycosyl donor (Figure 34b, Step 1) (i.e of opposite anomeric configuration to the natural hydrolytic substrate) to an acceptor bound in the aglycon pocket. The enzyme, without nucleophilic activity, could not form the covalent glycosyl-enzyme intermediate, and so hydrolysis of the product did not take place (Figure 34b, Step 2).

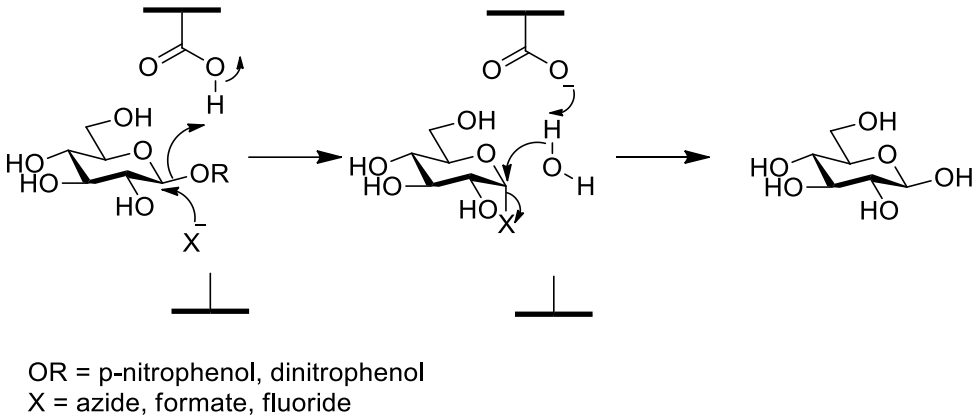
The groups of Planas[10] and Withers[1] investigated the use α-glycosyl fluorides as substrates that mimicked the glycosyl-enzyme intermediate. They observed that the glycosylation occurred with inversion of configuration to give a β-product, which was not subsequently hydrolysed by the enzyme. The rationale behind this observation was that the remaining acid/base residue catalysed the glycosylation step but the loss of the catalytic nucleophile eliminated the hydrolytic activity of the enzyme.

The use of such mutant enzymes, termed glycosynthases, is now well established in oligosaccharide synthesis, and the field undergoes continuous expansion with new classes of glycosidase mutants being developed.

4.1.4. Chemical rescue of mutant glycosidases

The theory on the use of chemical rescue for mutant glycosidases is described in Chapter 1 (Section 1.12). Briefly, glycosidases which have undergone a substitution of the catalytic nucleophile for a non-nucleophilic residue (such as glycine or alanine) have no (or in some cases, residual) activity towards their respective natural substrates. However, when the mutant glycosidase is presented with an activated glycosyl substrate, such as a *p*-nitrophenyl or dinitrophenyl glycoside, and a small charged

exogenous nucleophile, such as azide, formate, or fluoride, the hydrolytic activity can be restored (Scheme 39).



Scheme 39 - Chemical rescue of mutant glycosidases using exogenous nucleophiles.

4.1.5. Michaelis-Menten kinetics

Michaelis-Menten kinetics is one of the best representations of enzyme kinetics,[11] and serves to explain how an enzyme can cause kinetic rate enhancements of a reaction, and also shows how reaction rates depend on the concentration of enzyme and substrate. The Michaelis-Menten model for enzyme kinetics presumes a simple two-step reaction. The first step involves the binding of the substrate to the enzyme, which is then followed by catalysis (conversion of the substrate to product), and then finally the subsequent release of the product from the enzyme active site (Figure 35).

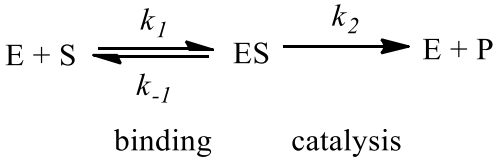


Figure 35 - General reaction scheme of an enzyme catalysed reaction. k_1 represents the rate constant for the binding of the substrate (S) to the enzyme (E) to form the enzyme-substrate complex (ES); k_{-1} represents the rate of dissociation of the (ES) complex to free (E) and (S); k_2 is the rate determining step and represents the catalysis reaction producing the reaction product (P) and regenerating free (E).

Based on several assumptions, explained below, the Michaelis-Menten model takes the form of an equation which describes the rate of enzymatic reactions by relating reaction rate (v) to substrate concentration [*S*]. The formula is given by:

$$v = \frac{d[P]}{dt} = \frac{V_{max}[S]}{K_M + [S]}$$

Here, V_{max} represents the maximum rate achieved when the active site of all the enzymes in the system are occupied. The ratio of constants $(k_{-1} + k_2)/k_1$ is itself a constant and is defined as the Michaelis constant (K_M):

$$K_M = \frac{k_{-1} + k_2}{k_1}$$

In cases where k_2 is much smaller than k_{-1} (i.e. where ES complex dissociation to E + S is faster than product formation), K_M is an inverse measure of the substrate's affinity for the enzyme (i.e. a small K_M indicates high affinity, meaning that the rate will approach V_{max} faster, and vice-versa). When k_2 is larger than k_{-1}, K_M values for the enzyme are meaningless because the Michaelis constant, no longer a measure of substrate affinity, is lower than the dissociation constant $K_s = k_{-1}/k_1$.[12] In such cases, it is important to establish whether the glycosylation (k_1) or deglycosylation (k_2) step is rate limiting. Sinnott and Souchard reported a Bronsted analysis of leaving group ability of the donor substrate and the subsequent effect on the rate of reaction.[13] They found that the deglycosylation step (k_2) was rate limiting when donors with good leaving

groups were used as substrates. Conversely, k_1 was rate limiting in cases where donors with poor leaving groups (i.e. leaving groups with a high pK_a) were used.

There are several underlying approximations involved in the derivation of the Michaelis-Menten equation:

1) The binding step (k_1) is fast, allowing the reaction to quickly reach equilibrium ratios of [E], [S], and [ES]. The catalytic step (k_2) is also slower, thus rate limiting.
2) At early time points, where initial velocity (V_0) is measured, [P] is approximately equal to zero; since if we observe V_0 for the reaction immediately after E and S are mixed in the absence of P, the rate of any back reaction, is negligible because its rate will be proportional to [P], and [P] is essentially 0.
3) ES immediately comes to a steady state, so [ES] is constant (throughout the measurement of initial velocity) – this is the steady-state approximation and is an important part of the Michaelis-Menten model.
4) Since [S] is significantly greater than [E] at the start of the reaction, the fraction of S that binds to E is negligible, and so [S] effectively remains constant at early time points.
5) There are only two forms of the enzyme, either free (E) or bound to the substrate (ES). Therefore, the total [E] is the sum of the free and substrate-bound concentrations.

An enzyme kinetics assay involves performing a series of experiments at varying substrate concentrations and monitoring the absorbance of a UV-active chromophore, such as *p*-nitrophenol, during the initial rate phase. The absorbance values are converted

to initial rate (v_0) values by application of the Beer-Lambert law, which relates the attenuation of light to the properties of the material through which the light is travelling, and is given by the equation:

$$A = \varepsilon c l$$

where the absorbance (A) is directly proportional to the concentration (c), path length (l), and the extinction coefficient (ε) of the chemical species being observed.

By plotting the initial rates against [S], and using non-linear regression of the Michaelis-Menten equation (often by means of software, such as GraFit or Prism), the values for the parameters V_{max} and K_M can be obtained (Figure 36). Other parameters such as k_{cat} and k_{cat}/K_M can be calculated from the data.

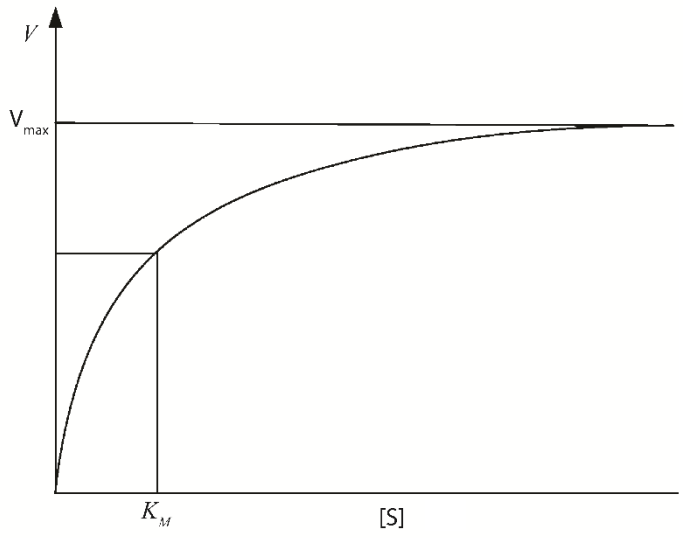


Figure 36 – A Michaelis-Menten plot.

k_{cat} is the turnover number of an enzyme, and is a measure of its maximal catalytic activity. k_{cat} is defined as the number of substrate molecules converted into product per

enzyme molecule per unit time when the enzyme is saturated with substrate, and can be calculated from a standard enzyme kinetics assay once V_{max} has been obtained, providing the total concentration of enzyme, $[E_T]$, is known. At saturating [S], $v = V_{max} = k_2[E_T]$. Thus,

$$k_2 = \frac{V_{max}}{[E_T]} = k_{cat}$$

Another parameter, k_{cat}/K_M, is the pseudo second-order rate constant for the reaction of E and S to form product, and a measure of the efficiency of Michaelis-Menten-type enzymes. The derivation can be seen when the conditions below are applied to the Michaelis-Menten equation:

$$v = \frac{V_{max}[S]}{K_M + [S]}$$

As $V_{max} = k_{cat}[E_T]$, where $[E_T]$ is the total amount of enzyme in the system:

$$v = \frac{k_{cat}[E_T][S]}{K_M + [S]}$$

When $[S] << K_M$, the concentration of free enzyme, [E], is approximately equal to $[E_T]$, so that:

$$v = \left(\frac{k_{cat}}{K_M}\right)[E][S]$$

Because K_M is inversely proportional to the affinity of the enzyme for its substrate, and k_{cat} is directly proportional to the kinetic efficiency of the enzyme, k_{cat}/K_M provides an index of the catalytic efficiency of an enzyme operating at substrate concentrations substantially below saturation amounts.

Due to the hyperbolic shape of ν versus [S] plots (Figure 36), V_{max} can only be estimated from an extrapolation of the asymptotic approach of ν to some limiting value as [S] increases indefinitely. Linear plots can be derived from the Michaelis-Menten equations.

One such plot is known as the Lineweaver-Burk double-reciprocal plot. Taking the reciprocal of both sides of the Michaelis-Menten equation yields:

$$\frac{1}{\mathrm{v}} = \left(\frac{\mathrm{K_M}}{\mathrm{V_{max}}}\right)\left(\frac{1}{[\mathrm{S}]}\right) + \frac{1}{\mathrm{V_{max}}}$$

This equation conforms to $y = mx + c$, where $y = 1/v$; m, the slope is K_M/V_{max}; $x = 1/[S]$; and $c = 1/V_{max}$. A plot of $1/\nu$ versus $1/[S]$ gives a straight line whose x-intercept is $1/V_{max}$, and slope is K_M/V_{max} (Figure 37).

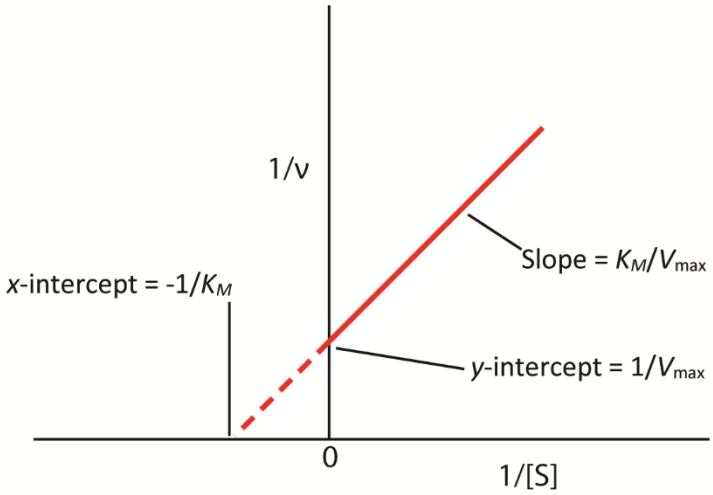


Figure 37 - The Lineweaver-Burk double-reciprocal plot, depicting extrapolations that allow the determination of the *x*-and *y*-intercepts and slope.

The expression and purification of engBL from *Bifidobacterium longum* and engEF from *Enterococcus faecalis*, as well as their mutagenesis, is detailed in Chapter 2. The kinetic parameters of the WT enzymes have been well characterised. However mutants of engBL and engEF enzyme have not. Therefore, it was necessary to perform enzyme kinetics assays to determine the kinetic parameters of the mutants and to demonstrate the extent to which mutagenesis had affected the activity of the respective WT enzymes. The Michaelis-Menten kinetics of the WT and mutant glycosidases investigated in this thesis were obtained and are described later in the relevant sections.

4.2. Objectives

Fujita *et al.*[9] and Ito *et al.*[14] demonstrated that both engBL and engEF, could catalyse the hydrolysis of the disaccharide Galβ1,3GalNAc (core 1) from mucin-type glycoproteins. EngBL could also catalyse the glycosylation of Galβ1,3GalNAc to various peptides containing serine or threonine residues.[15] Moreover, both enzymes could catalyse the hydrolysis Galβ1,3GalNAc-α-pNP (core 1-pNP), however, only engEF could hydrolyse both core 1-pNP and GlcNAcβ1,3GalNAc-α-pNP (core 3-

pNP), demonstrating that this enzyme possesses a broader substrate specificity.[14] These observations stimulated our interest and led us to investigate whether it was possible to use engBL and engEF to chemoenzymatically access *O*-glycan glycopeptides by using the glycosylation potential of the family GH101 endo-α-*N*-acetylgalactosaminidases. Furthermore, we were interested in performing mutageneses of the family GH101 enzymes to generate glycosynthase enzymes with improved glycosylation abilities. To date, the mutageneses of these two enzymes to generate glycosynthases has not been reported.

The expression and purification of engBL[15] and engEF[14], as well as their mutagenesis, is described in Chapter 2. Chapter 3 details the synthesis of various activated *O*-glycan donors as substrates for engBL and engEF. This chapter outlines the kinetic analyses that were performed on the WT and mutant GH101 glycosidases. The substrate specificity of the WT and mutant variants of engBL and engEF using glycosyl acetates (see Chapter 5) was investigated to determine whether these could be better donors for the glycosylation reaction compared to more common *p*-nitrophenyl glycoside substrates.

Investigation of the kinetic parameters of the WT and mutant engBL and engEF enzymes, using Michaelis-Menten steady-state approximations, is also detailed. Chemical rescue experiments were also performed using various nucleophiles to determine whether the activity of the mutants could be recovered. Finally, the glycosylation of serine by a variety of *O*-glycan donors using engBL, engEF, and their respective mutants was investigated.

4.3. EngBL from *Bifidobacterium longum*

4.3.1. The hydrolytic activity and Michaelis-Menten kinetics of WT engBL

Michaelis-Menten steady-state approximations were utilised to determine the kinetic parameters of engBL[16] for the hydrolysis of Galβ1,3GalNAc-α-pNP (core 1-pNP) (Figure 38). A continuous assay was used whereby the release of *p*-nitrophenol was directly monitored by UV spectroscopy at 235 nm (ε_{235nm} 1176 L mol^{-1} cm^{-1}). The details of the assay are described in Chapter 6.

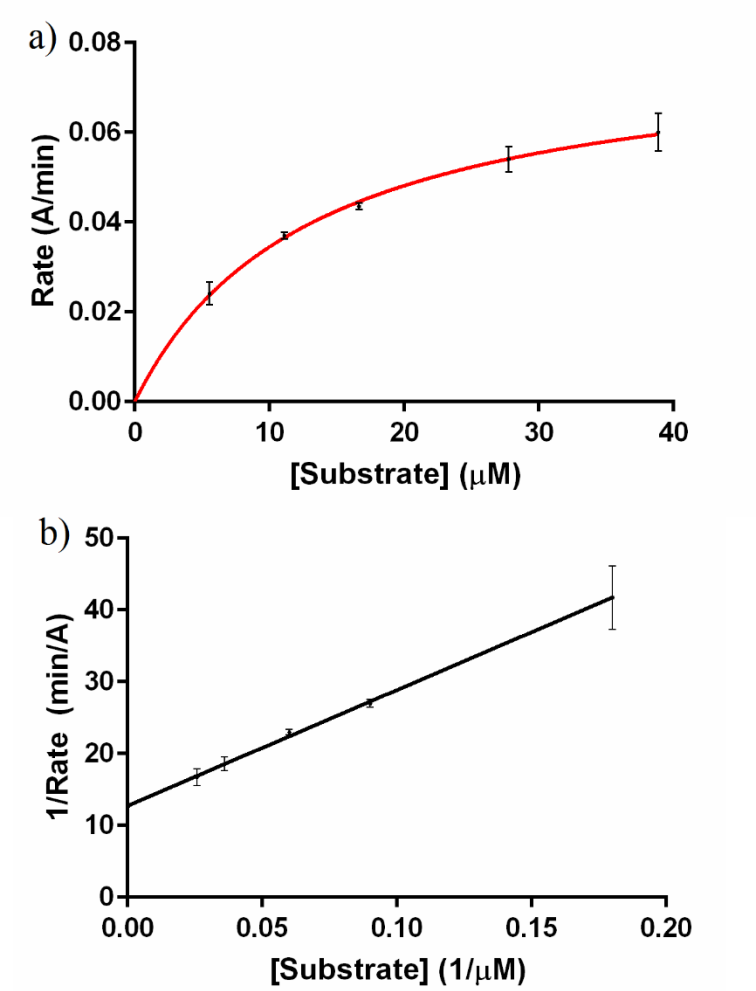


Figure 38 – a) Michaelis-Menten curve for WT engBL with Galβ1,3GalNAc-α-pNP (core 1-pNP). Data points are shown as black circles and a line of best fit in red; b) A Lineweaver-Burk plot.

The values for k_{cat} and K_M of Galβ1,3GalNAc-α-pNP for WT engBL were calculated to be 9.62 s^{-1} and 13 μM, respectively, varying slightly from the values of k_{cat} = 17.8 s^{-1} and K_M = 21.8 μM.[16] The differences in values attained in the previous studies are likely a result of variation in methodology and conditions used in the analysis.

The kinetics of the hydrolysis of the activated disaccharide, Galβ1,3GalNAc-α-OAc (core 1-α-OAc) were also investigated. Zinin *et al.* had found that glycosyl acetates were a better donor for β-galactosidase from *Penicillium sp.*[17] where the catalytic activity (k_{cat}) of the enzyme for the glycosyl acetate substrate was three orders of magnitude greater than that for the *p*-nitrophenyl glycoside.

Using methods published by Parsons *et al.*[18], Michaelis-Menten analysis of WT engBL revealed that for the hydrolysis of Galβ1,3GalNAc-α-OAc (core 1-α-OAc) the k_{cat} and apparent K_M were 26.7 s^{-1} and 23 μM, respectively (Figure 39).

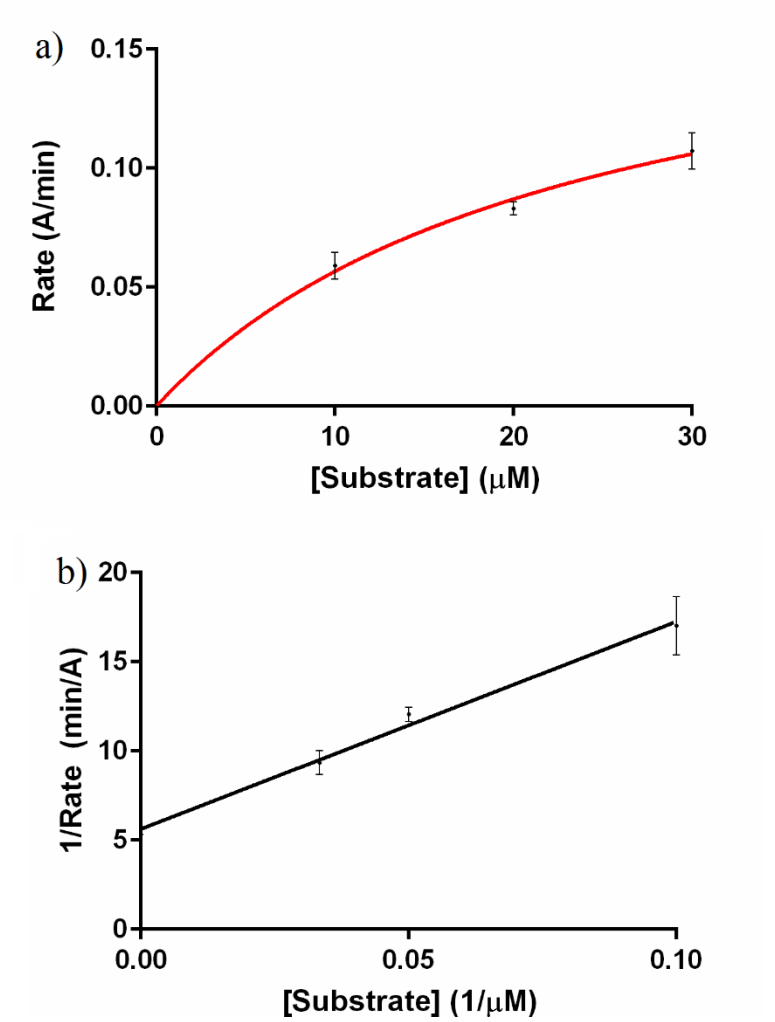


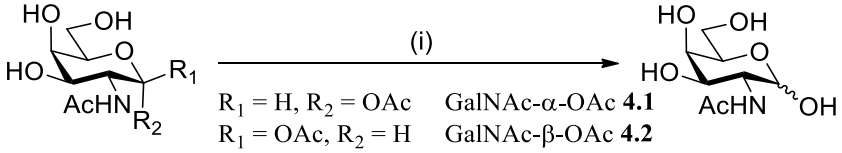
Figure 39 – a) Michaelis-Menten curve for WT engBL with Galβ1,3GalNAc-α-OAc (core 1-OAc). Data points are shown as black circles and a line of best fit in red. b) A Lineweaver-Burk Plot.

The use of core 1-α-OAc led to higher catalytic activity of the enzyme due to higher lability of the anomeric acetate as compared to the *p*-nitrophenyl glycoside. In terms of the kinetic scheme presented here, the data observed indicates that core 1-α-OAc is a better substrate for engBL (k_{cat}/K_M 1161 mM^{-1} s^{-1}) than core 1-α-pNP (k_{cat}/K_M 740 mM^{-1} s^{-1}), with the enzyme displaying almost two times the activity of that of the *p*-nitrophenyl glycoside. The K_M values for the pNP disaccharide (21.8 μM) and core 1-α-OAc (23 μM) were almost identical, indicating that the affinity of engBL for both substrates is essentially the same.

An investigation was made to determine whether WT engBL could hydrolyse the monosaccharide, GalNAc-OAc. The original research published by Fujita *et al*. indicated that engBL was not capable of hydrolysing the monosaccharide, GalNAc-α-pNP.[9]

To this end, both anomers of GalNAc-OAc were individually incubated with engBL (226 mU) for 16 h, and then directly freeze-dried and analysed by ^{1}H NMR. The International Union of Biochemistry (IUB) defines a standard unit (U) of enzyme activity as the amount of enzyme which catalyses the transformation of 1 µmol of the substrate per minute under standard conditions.[19] The amount of engBL required for one unit of activity was calculated in Chapter 2 to be 0.31 µg.

Non-enzymatically catalysed hydrolysis of GalNAc-OAc was also measured by incubating the respective anomers in buffer without the addition of enzyme.



Enzyme	GalNAc-α-OAc 4.1	GalNAc-β-OAc 4.2
No enzyme	22 %	45 %
engBL	48 %	56 %

Table 8 - Hydrolysis of monosaccharide donors **4.1** and **4.2** by engBL. Conditions: i) engBL from *Bifidobacterium longum* (226 mU), 25 mM NaOAc, pH 6.0, 37 ºC. After 16 h, yields were calculated by ^{1}H NMR by comparing the integrals of H-1 and $COCH_3$ of the respective sugars.

After 16 h, ^{1}H NMR analysis of the control (no enzyme) indicated that 45 % of the β-anomer **4.2** was hydrolysed compared to 22 % of that of the α-anomer **4.1**. The differences in the amount of hydrolysed starting material suggests that the mechanism of hydrolysis involves cleavage of the anomeric bond (Figure 40a, **1**), and not the C-O bond of the ester by an ester hydrolysis mechanism with water (Figure 40a, **2**). If the ester linkage was cleaved, a substantial difference in the amount of hydrolysis of the two different anomers should not be observed.

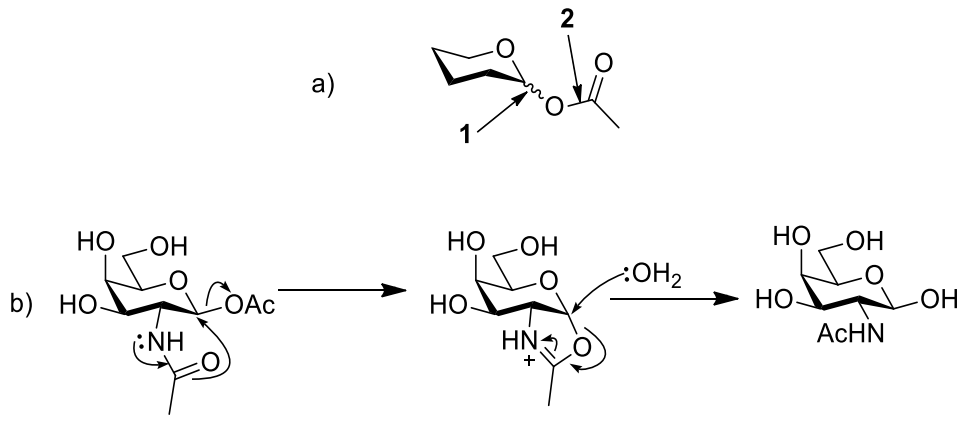


Figure 40 – Mechanism of spontaneous hydrolysis of GalNAc-β-OAc **4.2** in water. a) cleavage of the anomeric linkage **1** or ester linkage **2**; b) mechanism of hydrolysis *via* anchimeric assistance by the 2-acetamido group.

The rate of hydrolysis of the β-anomer **4.2** of GalNAc-OAc is likely accelerated by the presence of the acetamido functionality at C-2 which is suitably positioned to provide anchimeric assistance in the departure of the β-anomeric group (Figure 40b).

Next, both anomers of GalNAc-OAc were separately incubated with engBL. After 16 h, the α-anomer **4.1** was found to have hydrolysed by 48 %, whereas the β-anomer **4.2** was hydrolysed by 56 % (Table 8). Preliminary analysis of the reaction mixture containing the GalNAc-β-OAc **4.2** therefore suggests that the enzyme catalyses the

hydrolysis of the β-anomer, albeit very slowly. The ability of glycosidases to binding and performing catalysis on substrates of the 'wrong' configuration (i.e. opposite configuration to that of the natural substrate) is well documented for mutant glycosidases.[20] Therefore, it is possible that WT glycosidases can also bind the 'wrong' anomer of GalNAc-OAc (i.e. the β-anomer). However in this case, the amount of hydrolysis of GalNAc-β-OAc **4.2** was low. The low hydrolytic rate could arguably be due to steric hindrance by the nucleophilic residue (D789) which may interfere with the ability of GalNAc-β-OAc to fit into the binding pocket.

The observation that engBL bound and hydrolysed GalNAc-α-OAc **4.1** was also interesting since the initial investigations of engBL substrate specificity indicated that GalNAc-α-pNP was not a substrate for the enzyme.[9] The ability of engBL to catalyse the hydrolysis of GalNAc-α-OAc **4.1** is likely due to the increased lability of the anomeric acetate (pK_a ~ 4.76). Glycosyl acetates have previously been found to be better substrates for glycosylation reactions as compared to their *p*-nitrophenyl counterparts[17], and, for example, have increase the catalytic activity of the enzyme by up to 100-fold.

4.3.2. Chemical rescue and Michaelis-Menten kinetics of engBL mutants

A substitution of the catalytic nucleophile (D789) of engBL for residues glycine, alanine, and serine respectively was expected to affect the hydrolytic activity of the enzyme. To determine the magnitude of the effect of the D789 point mutations, the kinetic parameters for the mutant engBL enzymes were determined for the hydrolysis of Galβ1,3GalNAc-α-pNP (core 1-pNP) using Michaelis-Menten steady-state

approximations, and a continuous assay to monitor the release of *p*-nitrophenol by UV spectroscopy at 235 nm (ε_{235nm} 1176 L mol^{-1} cm^{-1}) (Figure 41).

As expected, the mutagenesis of engBL led to a significant decrease in the catalytic activity in the enzyme (Table 9). The turnover number (k_{cat}) of the D789G/A/S mutants, which represents the number of times each enzyme site converts substrate to product per unit time, all decreased by over 200-fold compared to that of WT engBL when incubated with Galβ1,3GalNAc-α-pNP (core 1-pNP). The highest substrate turnover was observed for the D789S mutant (Table 9, Entry 4) with a k_{cat} value of 0.69 s^{-1}, which is approximately 15 times lower than the WT enzyme. However, K_M increased 100-fold to 1.3 mM compared to that of the WT enzyme indicating that the substitution may have had a negative effect on the ability of the enzyme to bind the substrate. The activity of the D789G (Table 9, Entry 2) and D789A (Table 9, Entry 3) mutants were also significantly reduced, with measured k_{cat} values of 0.54 s^{-1} and 0.019 s^{-1}, respectively. The rate decreases observed here for the engBL D789 mutants collectively suggested that the mutagenesis has affected an important residue in the active site that is involved in catalysis. It is well reported that mutageneses which target the nucleophilic residue of glycosidases also lead to significant decreases in enzymatic activity.[21,22]

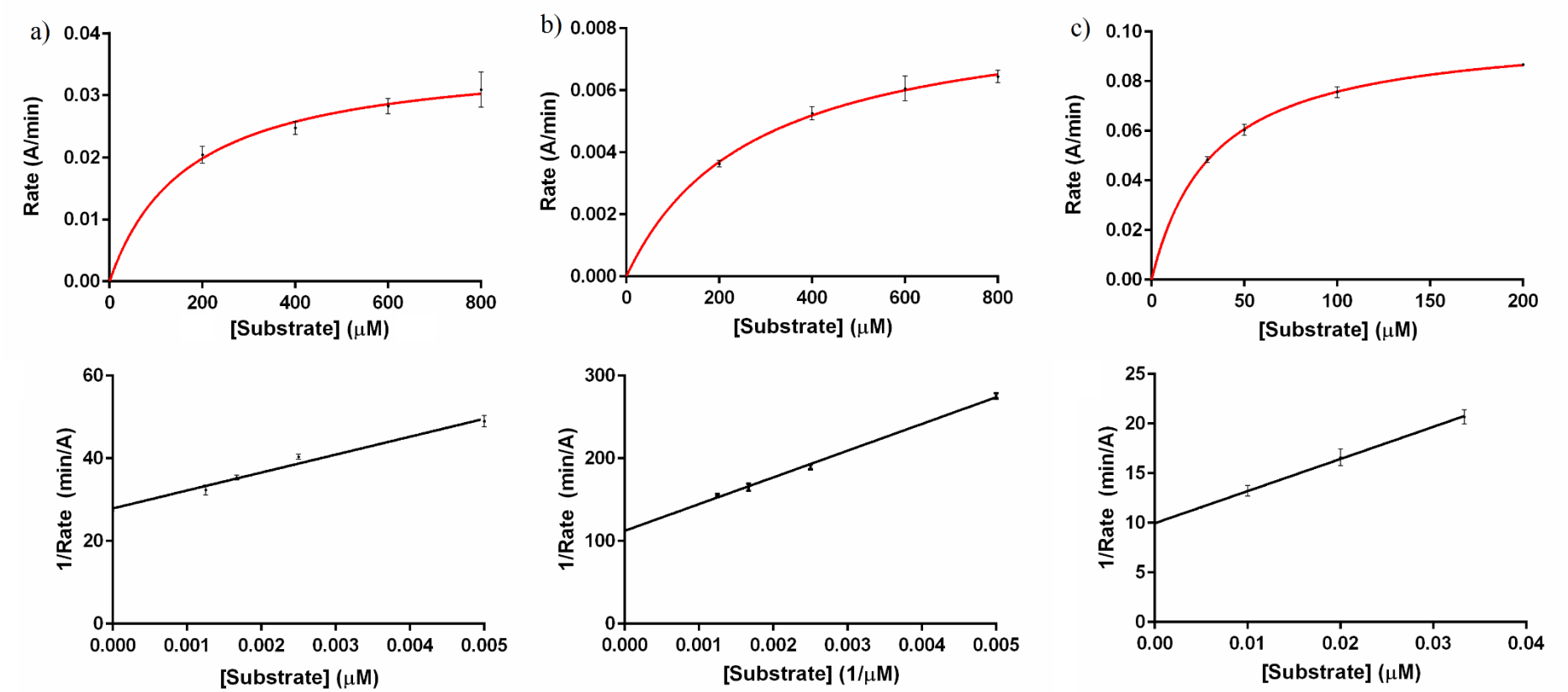
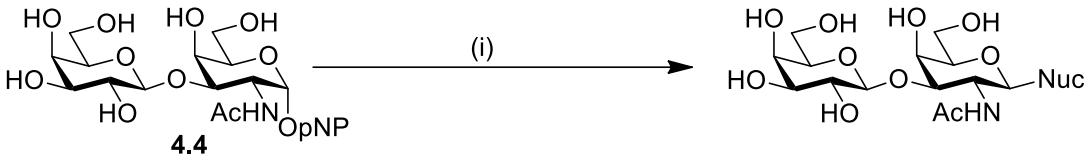


Figure 41 – (Top) Michaelis-Menten curves for a) engBL D789G; b) engBL D789A and; c) engBL D789S. Data points are shown as white circles and the line of best fit in red. (Bottom) A Lineweaver-Burk plot.



Entry	Enzyme	Nucleophile	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1} s^{-1})	Activity recovered[a]
1	WT	-	0.0131	9.62	738	-
2	D789G	-	0.170	0.543	3.20	0.434 %
3	D789A	-	0.272	0.0195	0.0714	0.00967 %
4	D789S	-	1.34	0.689	0.514	0.0696 %
5	D789G	100 mM NaN_3	0.0498	0.647	13.0	1.76 %
6	D789A	100 mM NaN_3	0.322	0.0347	0.108	0.0146 %
7	D789S	100 mM NaN_3	0.0321	0.244	7.61	1.03 %
8	D789G	200 mM NaBr	0.0616	0.385	6.25	0.847 %
9	D789A	200 mM NaBr	0.0481	0.00919	0.191	0.0259 %
10	D789S	200 mM NaBr	0.113	0.199	1.77	0.240 %
11	D789G	1 M NaHCOO	0.0486	0.435	8.96	1.21 %
12	D789A	1 M NaHCOO	0.0242	0.0123	0.508	0.0688 %
13	D789S	1 M NaHCOO	0.0307	0.102	3.32	0.450 %

Table 9 - Kinetics and chemical rescue of engBL D789G/A/S using various nucleophiles. Conditions: engBL D789G/A/S, nucleophile, 25 mM NaOAc buffer, pH 6.0, 37 °C. Chemical reactivation experiments of the inactive nucleophilic variants were also performed using NaN_3, NaBr, and NaHCOO as external nucleophiles. [a] activity recovered relative to WT engBL.

Chemical rescue experiments were performed to determine whether the activity of the mutants could be recovered in the presence of an exogenous nucleophile. As mentioned in Chapter 1, due to the absence of the catalytic nucleophile, the mutants should lose the ability to form the enzyme-substrate intermediate, therefore it was expected that the

engBL mutants would be inactive against the pNP disaccharide. However, in the presence of a small exogenous nucleophile such as azide, the activity of the enzyme should be restored.[23]

A variety of concentrations of the respective exogenous nucleophile were investigated to determine which, if any, concentrations would restore activity to the mutant enzymes. These analyses indicated 100 mM, 200 mM, and 1 M were appropriate for NaN_3, NaBr, and NaHCOO, respectively.

As shown in Table 9, chemical rescue of the mutants by the added various nucleophiles was not very effective. The best rescue was observed with the D789G mutant (Table 9, Entry 5), which recovered 1.8 % of its original catalytic efficiency (calculated by taking 100 % the k_{cat}/K_M of the WT enzyme (Entry 1)), when 100 mM NaN_3 was used as a nucleophile. The D789S (Entry 7) and D789A (Entry 6) mutants recovered low percentage activities, namely 1 % and 0.02 % respectively, of the WT enzyme.

Incubation of the mutants with alternative nucleophiles did not lead to greater recoveries of catalytic activity. The best activity was recovered when 200 mM of NaBr was used to rescue the D789G mutant (0.9 %, Entry 8), followed by D789A (Entry 9) and D789S (Entry 10) mutants (0.026 and 0.24 %, recovered activity respectively). The greater recovery observed for the D789G mutant could be due to the size of the cavity created upon removal of the carboxylate, allowing room for the large bromide ion to occupy the enlarged binding site (Figure 42).

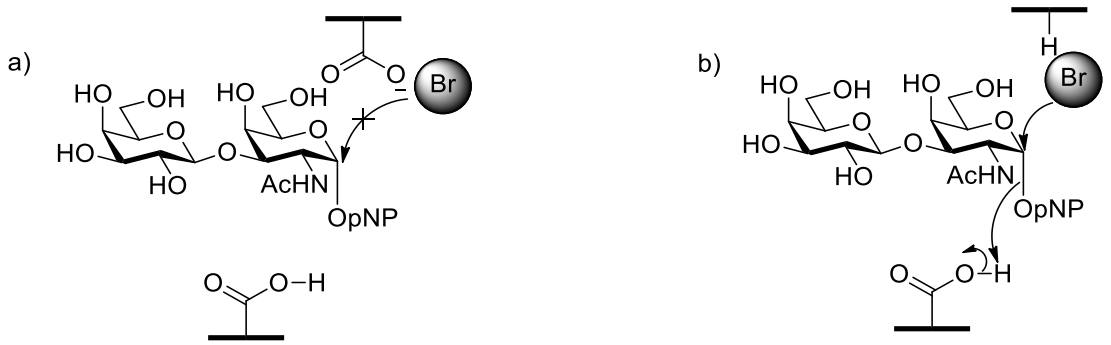
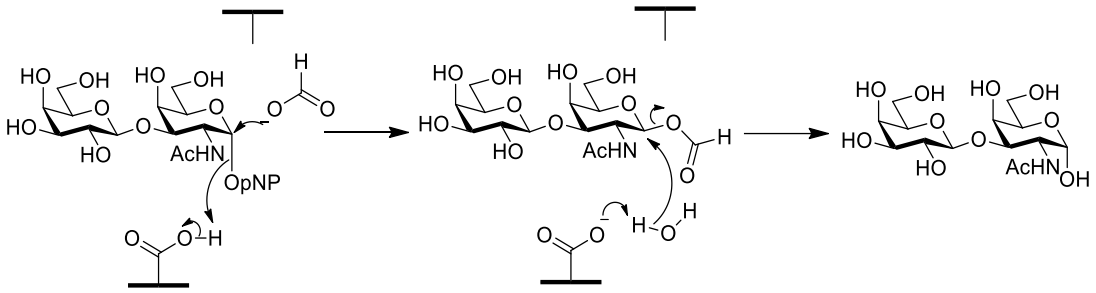


Figure 42 - Chemical rescue with bromide. a) Steric hindrance by the catalytic nucleophile means that the bromide nucleophile cannot access the binding site for chemical rescue; b) Substitution of the catalytic nucleophile for a smaller residue forms a large pocket which fits the exogenous bromide.

Finally, chemical rescue with sodium formate was investigated. It is known that exogenous formate can act as a mimic of the carboxylic acid of the enzyme and so may assist in the catalysis (Scheme 40).[24] In this case, the activated α-glycosyl donor should result in the *in situ* generation of β-glycosyl-intermediate which can be used by the glycosynthase.

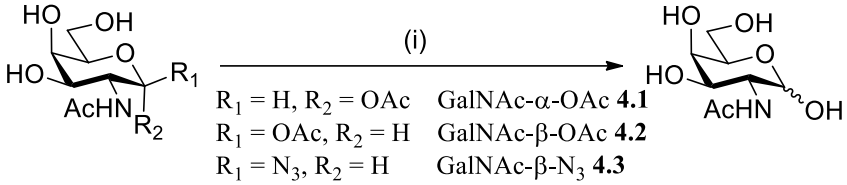
Similar decreases in rate constants have also been observed in reported chemical rescue experiments. For example, Hommalai *et al.* found that mutagenesis of rice β-glucosidase (BGlu1) affected the activity of the enzyme.[22] Chemical rescue experiments of the E414A and E414S mutants led to only modest increases in activity (up to 0.229 % and 0.0145 % for the E414A and E414S mutants, respectively. Hommalai *et al.* suggested that the bulkier hydroxymethyl of serine and the methyl group of alanine might hinder the attack of the nucleophile on the substrate. Cobucci-Ponzano *et al.* have also seen similar decreases in their chemical rescue experiments of mutants of the α-L-fucosidase from *Sulfolobus solfataricus* (SSα-Fuc) using formate and azide as external nucleophiles.[21]



Scheme 40 - Chemical rescue of mutant glycosidase using sodium formate.

Chemical rescue of the mutant glycosidases with sodium formate was also poor. Again, engBL D789G showed the greatest recovery (1.2 %, Entry 11), followed by the D789S (Entry 13), and D789A (Entry 12) mutants.

As with WT engBL, investigations into the hydrolytic activity of the engBL mutants were carried out using monosaccharide donors (Table 10).



Enzyme	Mutant	GalNAc-α-OAc **4.1**	GalNAc-β-OAc **4.2**	GalNAc-β-N_3 **4.3**
No enzyme	-	22 %	48 %	0 %
engBL	D789G	22 %	67 %	0 %
	D789A	23 %	69 %	0 %
	D789S	23 %	68 %	0 %

Table 10 - Hydrolysis of monosaccharide donors (100 nmol) by engBL mutants. Conditions: engBL D789G/A/S, 25 mM NaOAc buffer, pH 6.0, 16 h, 37 °C. After 16 h, yields were calculated by ^{1}H NMR by comparing the integrals of H-1 and $COCH_3$ of the starting material and hydrolysed sugars.

The spontaneous hydrolysis of GalNAc-α-OAc **4.1** was measured by incubating the glycoside in the reaction conditions without enzyme. After 16 h, it was found that 22 % of the starting material had hydrolysed to produce the hemiacetal.

Due to the absence of a catalytic nucleophile, it was expected that the engBL mutants would not be able to catalyse the hydrolysis of α-anomer **4.1** of GalNAc-OAc. After 16 h of incubation of GalNAc-α-OAc **4.1** with the respective mutants, analysis of the reaction mixture indeed indicated that only background hydrolysis was occurring, suggesting again that a catalytically important residue had been removed.

Hydrolysis of the β-anomer **4.2** of GalNAc-OAc did occur albeit very slowly, with a maximum yield of 69 % when the substrate was incubated with the D789A mutant. This result (i.e. that the mutant enzyme was capable of hydrolysis of the β-acetate) is consistent with the understanding that mutant glycosidases are able to bind activated substrates of the opposite stereochemical configuration to that of the natural substrate and catalyse glycosylation reactions.[1,10]

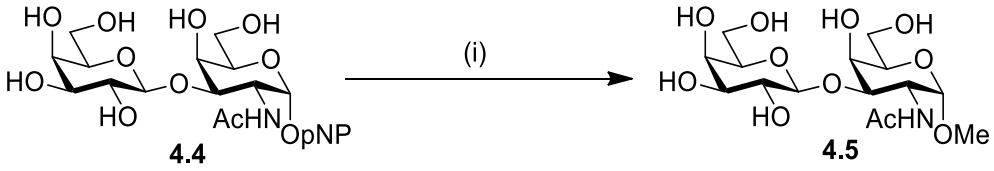
Cobucci-Ponzano *et al.* have observed that glycosyl azides can also be donors for glycosynthases.[21] Therefore, we wanted to determine whether GalNAc-β-N_3 **4.3** was a viable donor for our engBL mutant enzymes. To this end, GalNAc-β-N_3 **4.3** was also incubated with the engBL mutants. Disappointingly, ^{1}H NMR analysis showed that no hydrolysis of the starting material suggesting that glycosyl azides were not donors for the mutant engBL enzymes.[21]

4.3.3. Glycosylation reactions catalysed by engBL

4.3.3.1. Glycosylation of MeOH using WT engBL

The glycosylation activity of WT engBL was investigated. Firstly, the reported glycosylation of MeOH by Galβ1,3GalNAc-α-pNP **4.4** (core 1-pNP) was performed to determine whether the enzyme could perform the same reaction.

Core 1-pNP was incubated with engBL (226 mU) with 15 % v/v MeOH for 16 h, at which point, the reaction was quenched by addition of 1 M aqueous NaOH (Scheme 41). Analysis by mass spectrometry and RP-HPLC confirmed the formation of the expected methyl glycoside **4.5** in 49 % yield indicating that the expressed enzyme was active as described.[9]

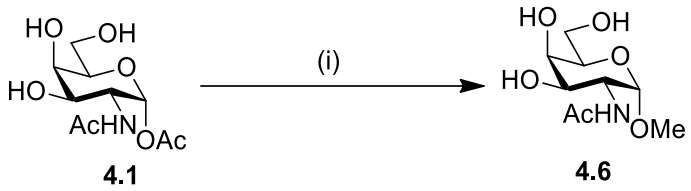


Scheme 41 - Glycosylation of MeOH by engBL. Conditions: engBL from *Bifidobacterium longum* (226 mU), 15 % MeOH, 25 mM NaOAc buffer, pH 6.0, 37 ˚C, 16 h, 49 %.

As described in Section 4.3.2., we observed that WT engBL could catalyse the hydrolysis of the monosaccharide, GalNAc-α-OAc **4.1**. Based on this result, we expected the enzyme should also catalyse glycosylation of MeOH by GalNAc-α-OAc **4.1**.

Therefore, GalNAc-α-OAc **4.1** was incubated with 15 % MeOH (v/v) and WT engBL (226 mU) for 16 h (Scheme 42). After 16 h, EngBL could catalyse the glycosylation of

MeOH with GalNAc-α-OAc **4.1**, showing 15 % yield of the desired methyl glycoside **4.6** and hydrolysis of the remaining substrate.

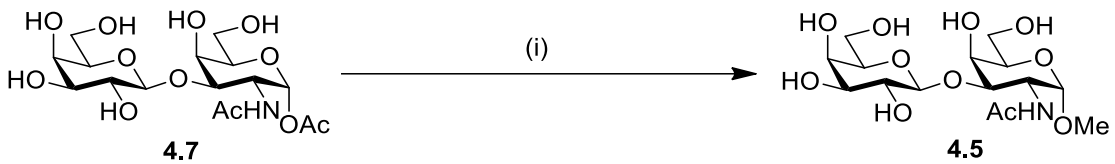


4.1 **4.6**

Scheme 42 - Glycosylation of MeOH to GalNAc-α-OAc **4.1** by engBL. Conditions: engBL from *Bifidobacterium longum* (226 mU), 15 % MeOH, 25 mM NaOAc buffer, pH 6.0, 37 ˚C, 16 h, 15 %.

Analysis of the control reaction, where no enzyme was added, showed that there was no methyl glycoside was formed, indicating that the enzyme was indeed responsible for catalysing formation of the methyl glycoside **4.6**.

It was found that WT engBL could also catalyse the glycosylation of MeOH by Galβ1,3GalNAc-α-OAc **4.7** (core 1-α-OAc) (Scheme 43). After 16 h, the desired methyl glycoside **4.5** was obtained in 38 % yield. Disappointingly, the yield obtained using the glycosyl acetate donor was lower than that which was reported for the corresponding *p*-nitrophenyl glycoside (49 %). Due to the lack of substrate, we could not optimise this yield any further.



4.7 **4.5**

Scheme 43 - Glycosylation of MeOH to core 1-α-OAc **4.7** by WT engBL. Conditions: engBL from *Bifidobacterium longum* (226 mU), 15 % MeOH (v/v), 25 mM NaOAc buffer, pH 6.0, 37 ˚C, 16 h, 38 %.

4.3.3.2. Glycosylation of MeOH by mutants of engBL

An attempt was made to glycosylate MeOH with activated β-glycosyl donors using the engBL mutant enzymes. The first set of reactions was performed using GalNAc-β-OAc **4.2** as the donor. As observed in Section 4.3.1, WT engBL was capable of catalysing the hydrolysis of the monosaccharide, GalNAc-α-OAc **4.1**. In theory, the mutant engBL enzymes may be able to bind substrates of the opposite configuration, in this case, GalNAc-β-OAc **4.2**, and catalyse the transfer of GalNAc to an appropriate acceptor to give the product with inverted stereochemistry (Figure 43).

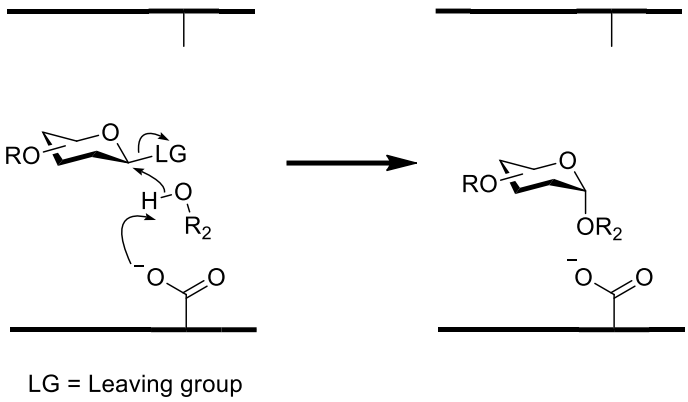
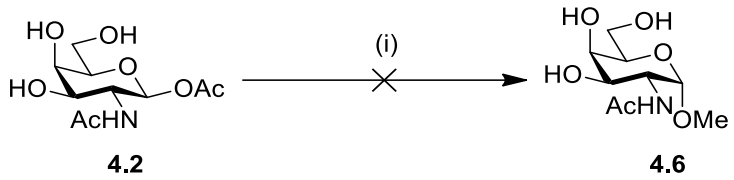


Figure 43 - Mechanism for the transfer of an acceptor by a GH101 glycosynthase.

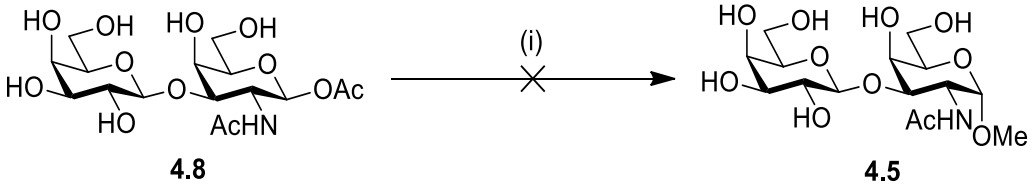
GalNAc-β-OAc **4.2** was incubated with 15 % MeOH (v/v), and the mutant enzymes engBL D789G/A/S. A control reaction was also performed where enzyme was not added.



4.2 **4.6**

Scheme 44 - Attempted glycosylation of MeOH to GalNAc-β-OAc **4.2** using engBL D789G/A/S. Conditions: engBL D789G/A/S (20 µL of a 14.3/9.4/8.6 mg/mL solution), 15 % MeOH, 25 mM NaOAc buffer, pH 6.0, 37 ˚C, 16 h.

To our disappointment, the formation of the methyl glycoside **4.6** was not observed for any of the mutant enzymes, by either RP-HPLC (product retention time confirmed by synthesis of the methyl glycoside by the chemical means) or t.l.c. analysis. Although this observation suggests that the mutants are unable to catalyse the transfer of MeOH to GalNAc, it was difficult at that point to conclude that the enzyme was completely inactive towards activated β-glycosyl donors since the monosaccharide is not the preferred substrate for the GH101 endo-α-*N*-acetylgalactosaminidases. A more appropriate substrate to glycosylate MeOH was the therefore Galβ1,3GalNAc (core 1) disaccharide.[9] To this end, Galβ1,3GalNAc-β-OAc **4.8** and MeOH were incubated with the engBL D789G/A/S mutants for 16 h (Scheme 45).



4.8 **4.5**

Scheme 45 - Attempted glycosylation of MeOH by core 1-β-OAc **4.8** using engBL D789G/A/S mutants. Conditions: engBL D789G/A/S (20 µL of a 14.3/9.4/8.6 mg/mL solution), 15 % MeOH, 25 mM NaOAc buffer, pH 6.0, 37 ˚C, 16 h.

After 16 h, RP-HPLC analysis of the reaction mixture indicated that the core 1-β-OAc **4.8** had completely hydrolysed, but without any product formation. Additionally, t.l.c.

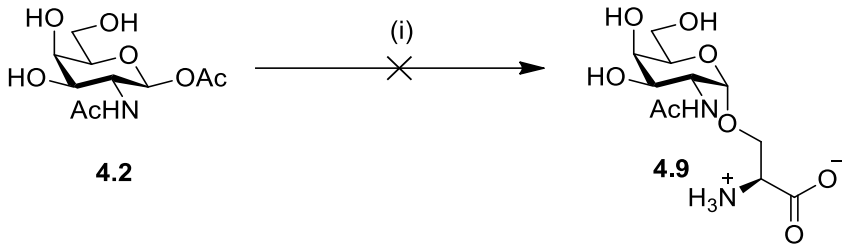
analysis of the reaction mixture at hourly intervals did not indicate any formation of the desired methyl glycoside.

To determine whether it was a substrate for the enzyme, core 1-β-OAc **4.8** was incubated in the presence and absence of engBL D789G for 16 h. Direct analysis of the respective reaction mixtures by RP-HPLC indicated that yields of the non-enzymatic-catalysed and enzymatic-catalysed reactions were essentially the same (87 % without enzyme, 84 % with enzyme).

In light of this disappointing result further investigations were abandoned.

4.3.3.3. Glycosylation of Ser-OH by mutant engBL

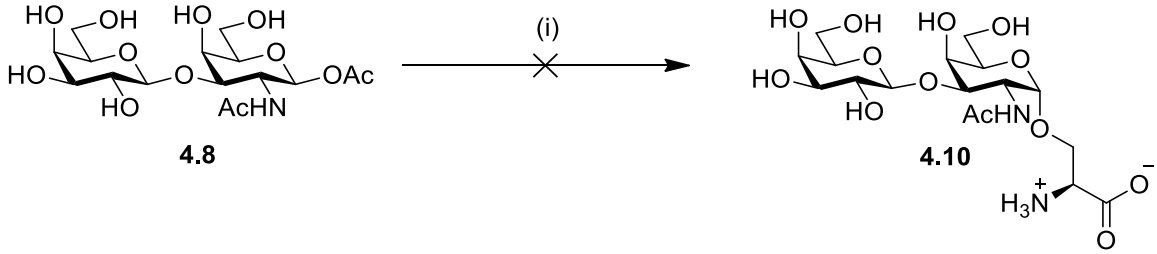
The aim of this project was to develop chemoenzymatic methodology for the synthesis of *O*-linked glycoconjugates. Therefore, as a prior step to peptides, it was appropriate to determine whether the engBL mutants could transfer single serine and threonine units to activated core *O*-glycan structures. Ashida *et al.* found that yields of glycosylation using serine as the acceptor were greater than when threonine was used.[15] Therefore, GalNAc-β-OAc **4.2** was incubated with 10 equivalents of Ser-OH and engBL D789G/A/S mutants to investigate whether the enzyme could catalyse the glycosylation of the unprotected amino acids (Scheme 46). A control reaction was performed where no enzyme was added.



Scheme 46 - Attempted glycosylation of Ser-OH to GalNAc-β-OAc **4.2** using engBL D789G/A/S mutants. Conditions: engBL D789G/A/S (20 µL of a 14.3/9.4/8.6 mg/mL solution), Ser-OH (10 equiv), 25 mM NaOAc buffer, pH 6.0, 37 ˚C, 16 h.

The reaction was monitored by t.l.c. and after 16 h, the starting material was found to be completely hydrolysed with no evidence of desired product formation. The absence of product was also confirmed with RP-HPLC and mass spectrometry.

To confirm the lack of any glycosynthase activity was not due to substrate preference, Galβ1,3GalNAc-β-OAc **4.8** (core 1-β-OAc) was also incubated with 10 equivalents of Ser-OH and the mutant enzymes.



Scheme 47 - Attempted glycosylation of Ser-OH to Galβ1,3GalNAc-β-OAc **4.8** using engBL D789G/A/S mutants. Conditions: engBL D789G/A/S (20 µL of a 14.3/9.4/8.6 mg/mL solution), Ser-OH (10 equiv), 25 mM NaOAc buffer, pH 6.0, 37 ˚C, 16 h.

After 16 h, RP-HPLC and t.l.c. analysis showed complete hydrolysis of core 1-β-OAc **4.8** indicating that the engBL mutants are incapable of catalysing the transfer of serine to the core 1-β-OAc **4.8** donor. Mass spectrometry did not indicate the presence of the desired glycosyl amino acid **4.10**.

4.3.4. Conclusions

This section details investigations into the properties of engBL and its mutant variants. Michaelis-Menten kinetics obtained for WT engBL determined k_{cat} and K_M for Galβ1,3GalNAc-α-pNP (core 1-pNP) were determined to be 9.62 s^{-1} and 13 μM, respectively. The kinetic parameters for the hydrolysis of disaccharide, Galβ1,3GalNAc-α-OAc (core 1-pNP), were found to be 26.7 s^{-1} and 23 μM for k_{cat} and K_M, respectively. These values were in line with the observation that glycosyl acetates are better donors for glycosidases than pNP glycosides.[17]

WT engBL could hydrolyse GalNAc-α-OAc, which was unexpected since the original investigations[9] had found that GalNAc-α-pNP was not a substrate for this enzyme. The glycosylation of methanol by both GalNAc-α-OAc and Galβ1,3GalNAc-α-OAc using WT engBL, in unprecedented reactions, were successful indicating that glycosyl acetates are viable substrates for the GH101 endo-α-*N*-galactosaminidases.

The activity and kinetic parameters of the three engBL mutants were also investigated. Chemical rescue experiments with azide, formate, and bromide resulted in low recovery of activity (< 2 %) in all cases when compared to the activity of WT engBL, suggesting that the enzyme active site had been perturbed by the mutagenesis. Disappointingly, the engBL mutants could not catalyse the glycosylation of either MeOH or Ser-OH by either GalNAc-β-OAc or Galβ1,3GalNAc-β-OAc. The underlying reasons for these observations are not clear at this point in time. Future computational and crystallographic studies may shed light on the impact of the point mutation on the active site of the enzyme.

4.4. EngEF from *Enterococcus faecalis*

4.4.1. Michaelis-Menten kinetics of WT engEF

Michaelis-Menten steady-state approximations were utilised to determine the kinetic parameters of WT engEF for the hydrolysis of Galβ1,3GalNAc-α-pNP (core 1-pNP). The hydrolysis of core 1-pNP by engEF was monitored by UV spectroscopy using a continuous assay at various concentrations of core 1-pNP. The release of *p*-nitrophenol was directly monitored at 235 nm (ε_{235nm} 1176 L mol^{-1} cm^{-1}). The values for K_M and k_{cat} for WT engEF were previously reported to be 59.6 μM and 49.9 s^{-1}, respectively.[14] For the engEF expressed for this work, the values for K_M and k_{cat} were found to be 53.5 μM and 76.4 s^{-1}, respectively, confirming that the protein expressed for this research to be similar in activity to the previously reported[14] enzyme (Figure 44).

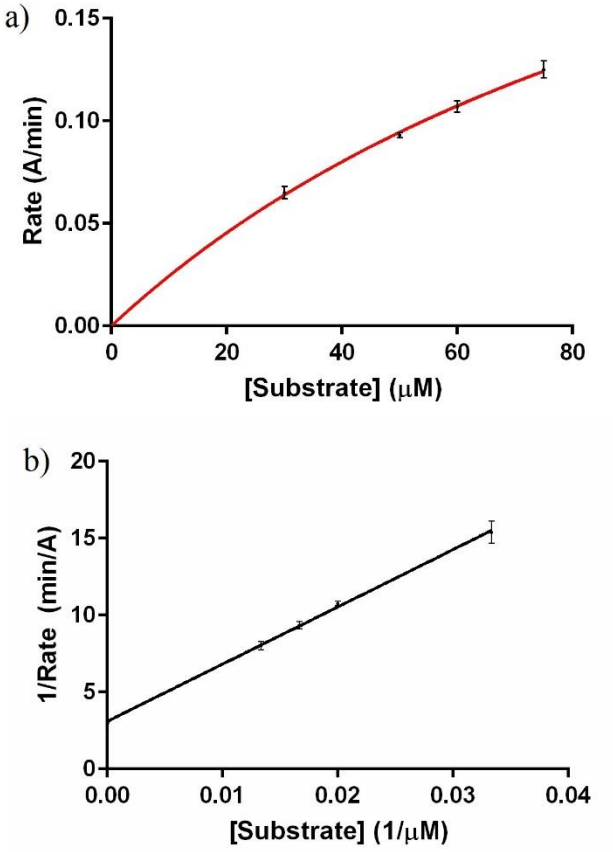


Figure 44 - a) Michaelis-Menten curve for WT engEF with Galβ1,3GalNAc-α-pNP (core 1-pNP). Data points are shown as black circles and a line of best fit in red. b) A Lineweaver-Burk plot.

The kinetic parameters of hydrolysis were also determined using Galβ1,3GalNAc-α-OAc (core 1-α-OAc) as the substrate. Using methods published by Parsons *et al.*[18], the values for K_M and k_{cat} were calculated to be 42 μM and 104 s^{-1}, respectively (Figure 45). As previously seen with WT engBL, the use of core 1-α-OAc led to higher catalytic activity of the enzyme due to higher lability of the anomeric acetate as compared to the *p*-nitrophenyl glycoside. This observation demonstrates that core 1-α-OAc is a better substrate for engEF (k_{cat}/K_M 2476 mM^{-1} s^{-1}) than core 1-α-pNP (k_{cat}/K_M 1428 mM^{-1} s^{-1}), with the enzyme displaying almost two times the activity of that of the *p*-nitrophenyl glycoside. The similar K_M values suggests that the affinity of engEF for both the core 1-α-pNP (53.5 μM) and core 1-α-OAc (42 μM) might be identical.

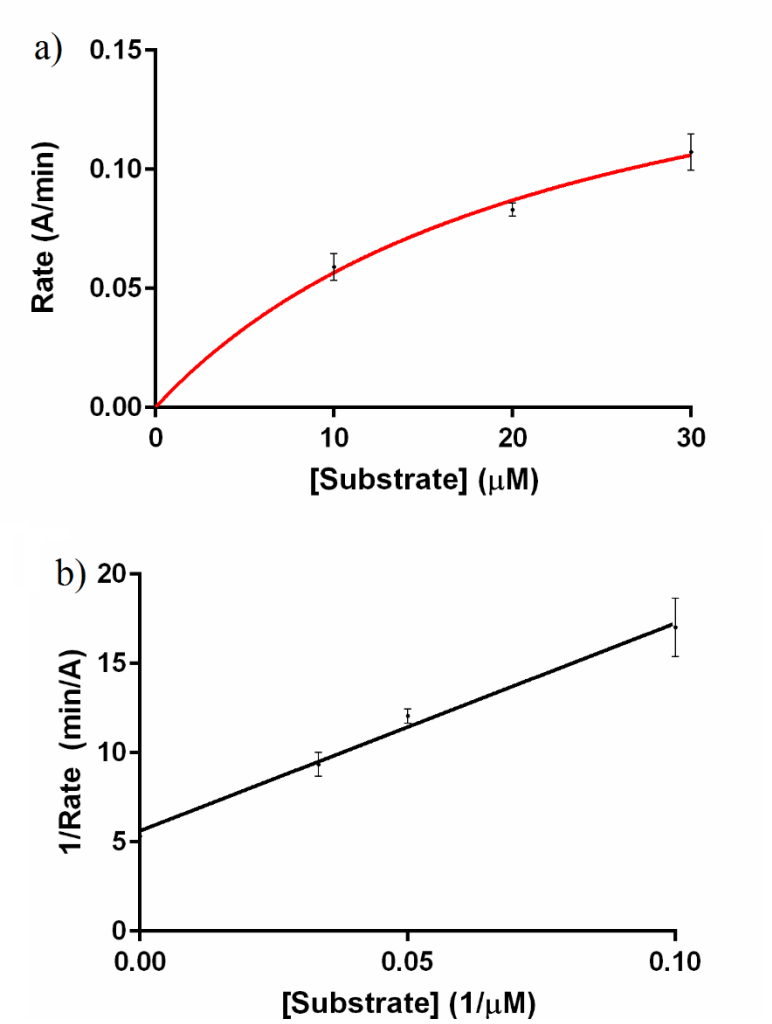
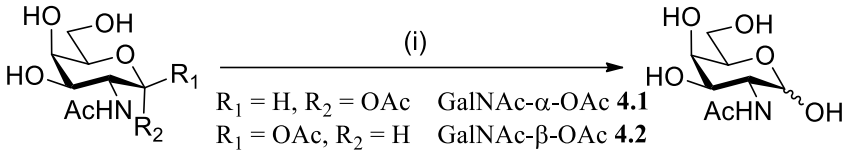


Figure 45 - a) Michaelis-Menten curve for WT engEF with Galβ1,3GalNAc-α-OAc (core 1-OAc). Data points are shown as white circles and a line of best fit in red. b) A Lineweaver-Burk plot.

The monosaccharide, GalNAc-OAc, was also tested as a substrate for hydrolysis by WT engEF. Both the α-anomer **4.1** and β-anomers **4.2** of GalNAc-OAc (100 nmol) were separately incubated with engEF (110 mU) for 16 h at 37 ºC, at which point the reaction mixture was lyophilised and analysed by ^{1}H NMR. Interestingly, WT engEF catalysed the hydrolysis of the α-anomer of GalNAc-OAc **4.1** with a greater efficiency compared to WT engBL, with 88 % hydrolysis observed after 16 h (Table 11).



Enzyme	GalNAc-α-OAc 4.1	GalNAc-β-OAc 4.2
No enzyme	22 %	45 %
engEF	88 %	56 %

Table 11 - Hydrolysis of GalNAc-OAc **4.1** and **4.2** (100 nmol) by WT engEF. Conditions: engEF from *Enterococcus faecalis* (110 mU), 25 mM NaOAc buffer, pH 6.0, 37 ˚C. After 16 h, yields were calculated by ^{1}H NMR by comparing the integrals of H-1 and $COCH_3$ of the respective sugars of the starting material and hydrolysed sugars.

Koutsioulis *et al.* reported that engEF could catalyse the hydrolysis of GalNAc-α-pNP; however the amount of *p*-nitrophenol released after 16 h was only 2.2 %, though the amount of enzyme used was not stated.[25] Goda *et al*. also reported the same reaction with engEF, and observed approximately 60 % hydrolysis of Galβ1,3GalNAc-α-pNP (core 1-pNP) when the substrate was incubated with 300 mU of enzyme for 16 h.

Incubation of GalNAc-β-OAc **4.2** with WT engEF (110 mU) did not result in a significant amount of hydrolysis, and only 11 % of the substrate was converted to the hemiacetal after 16 hrs.

4.4.2. The hydrolytic activity and Michaelis-Menten kinetics of mutant engEF

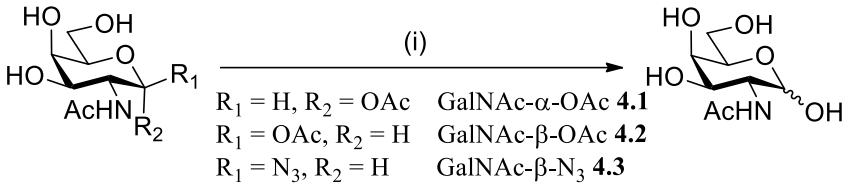
The mutant enzymes engEF D725G/A/S were incubated with Galβ1,3GalNAc-α-pNP **4.4** (core 1-pNP) to determine whether the mutations of the nucleophilic residue had affected the activity of the enzymes (Table 12). Chemical rescue experiments were also performed with azide, formate, and bromide as exogenous nucleophiles.

Entry	Enzyme	Nucleophile	k_M (mM)	K_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1} s^{-1})	Recovered activity[a]
1	WT	-	0.064	49	763	-
2	D789G	-	N.D.	N.D.	N.D.	N.D.
	D789A	-	N.D.	N.D.	N.D.	N.D.
3	D789S	-	N.D.	N.D.	N.D.	N.D.
4	D789G	100 mM NaN_3	N.D.	N.D.	N.D.	N.D.
5	D789A	100 mM NaN_3	N.D.	N.D.	N.D.	N.D.
6	D789S	100 mM NaN_3	N.D.	N.D.	N.D.	N.D.
7	D789G	200 mM NaBr	N.D.	N.D.	N.D.	N.D.
8	D789A	200 mM NaBr	N.D.	N.D.	N.D.	N.D.
9	D789S	200 mM NaBr	N.D.	N.D.	N.D.	N.D.
10	D789G	1 M NaHCOO	N.D.	N.D.	N.D.	N.D.
11	D789A	1 M NaHCOO	N.D.	N.D.	N.D.	N.D.
13	D789S	1 M NaHCOO	N.D.	N.D.	N.D.	N.D.

Table 12 - Kinetic parameters for the hydrolysis of core 1-pNP by engEF mutants. N.D. Not Detected. [a]Compared with the activity of WT engEF.

In contrast to the engBL mutants, the engEF mutants were all completely inactive when incubated with Galβ1,3GalNAc-α-pNP **4.4** (core 1-pNP) as monitored by UV spectroscopy. Moreover, the mutants could not be chemically rescued by azide, formate, or bromide (Table 12). As the engEF mutants were confirmed to be folded using circular dichroism spectroscopy (see Chapter 2), it was unclear which aspect of the protein had been affected by the mutagenesis.

The hydrolytic activity of the engEF mutants on both anomers of GalNAc-OAc (**4.1** and **4.2**) and GalNAc-β-N_3 **4.3** was examined (Table 13).



Enzyme	**Mutant**	**GalNAc-α-OAc 4.1**	**GalNAc-β-OAc 4.2**	**GalNAc-β-N_3 4.3**
No enzyme	-	22 %	48 %	0 %
engEF	D725G	23 %	50 %	0 %
	D725A	22 %	53 %	0 %
	D725S	23 %	53 %	0 %

Table 13 - Hydrolysis of GalNAc substrates using mutant engEF. Conditions: engEF D725G/A/S, 25 mM NaOAc buffer, pH 6.0, 37 ˚C. Hydrolysis after 16 h was measured by integral analysis of H-1 and $COCH_3$ of the respective anomers of GalNAc-OAc and GalNAc.

The non-enzymatic-catalysed rate of hydrolysis of the GalNAc-OAc substrates after 16 h was found to be 22 % and 48 % for the α-anomer **4.1** and β-anomer **4.2**, respectively. The engEF mutants were also incubated with GalNAc-β-OAc **4.2**. After 16 h, there was no appreciable amount of extra hydrolysis (i.e. only ~5 % difference from base hydrolysis) was observed for any of the mutant enzymes.

The hydrolysis of GalNAc-α-OAc **4.1** by the mutant engEF enzymes was also investigated. After 16 h, approximately 23 % of GalNAc-α-OAc **4.1** was hydrolysed in all cases, i.e. the same amount as the control, suggesting that the mutant ENGases did not catalyse the hydrolysis of GalNAc-α-OAc **4.1**. This result was expected since the

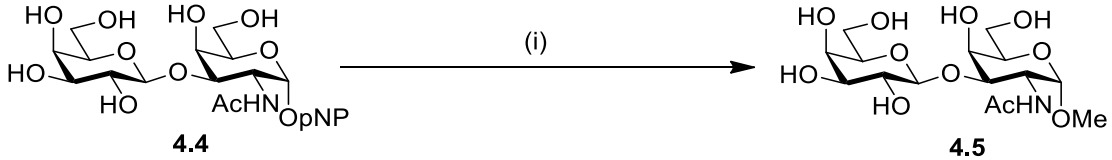
catalytic nucleophile has been removed, and so the enzyme is no longer capable of forming the glycosyl-enzyme intermediate required for hydrolysis.[1]

The engEF mutants were also incubated with GalNAc-β-N_3 **4.3**. However, after 16 h, there was no hydrolysis of the substrate indicating that glycosyl azides are not viable substrates for engEF.

4.4.3. Reactions catalysed by engEF

4.4.3.1. Glycosylation of MeOH using WT engEF

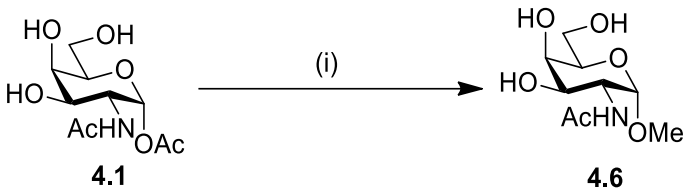
Goda *et al.* reported that WT engEF could catalyse the glycosylation of both MeOH and EtOH using Galβ1,3GalNAc-α-pNP **4.4** (core 1-pNP) as the donor.[14] We wanted to determine whether the engEF enzyme that we had expressed and purified in Chapter 2 could perform the same reaction. To this end, core 1-pNP **4.4** was incubated with engEF with 15 % MeOH (v/v) as described by Goda *et al* (Scheme 48).



Scheme 48 - Glycosylation of MeOH using Galβ1,3GalNAc-α-pNP **4.4** (core 1-pNP) as the donor. Conditions: i) engEF from *Enterococcus faecalis* (110 mU), 15 % v/v MeOH, 25 mM NaOAc buffer, pH 6.0, 37 ˚C, 16 h, 56 %.

RP-HPLC and mass spectrometric analysis confirmed the formation of the desired methyl glycoside **4.5** in 56 % yield, indicating that the enzyme expressed in this research was active as described.[14]

Our attention then turned to the glycosylation of MeOH with GalNAc-α-OAc **4.1**. As stated in Section 4.4.1, WT engEF was capable of hydrolysing the monosaccharide, and therefore we expected that the enzyme may be able to catalyse the glycosylation of MeOH. To this end, GalNAc-α-OAc **4.1** and WT engEF were incubated with 15 % MeOH (v/v) as the acceptor for 16 h (Scheme 49).

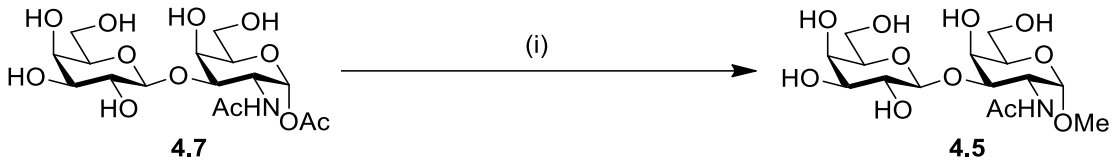


4.1 **4.6**

Scheme 49 - Glycosylation of MeOH to GalNAc-α-OAc **4.1** using WT engEF. Conditions: engEF from *Enterococcus faecalis* (111 mU), 15 % MeOH (v/v), 25 mM NaOAc buffer, pH 6.0, 37 ˚C, 16 h, 40 %.

Analysis of the reaction mixture after this time indicated that the enzyme had transferred MeOH to the donor with a 40 % yield to the methyl glycoside **4.6**, as observed by ^{1}H NMR. T.l.c. also indicated the formation of a major product of lower polarity as compared to the starting material **4.1**. The control reaction mixture (no enzyme) did not contain any methyl glycoside **4.6**. The higher yield observed for engEF-catalysed glycosylation of MeOH to GalNAc-α-OAc **4.1** compared to that with engBL could be due to the broader substrate specificity of the former enzyme.[25]

Finally, we wanted to determine whether WT engEF could also catalyse the glycosylation of MeOH by Galβ1,3GalNAc-α-OAc **4.7** (core 1-α-OAc) (Scheme 50). Core 1-α-OAc **4.7** was therefore incubated with 15 % MeOH (v/v) and WT engEF.



4.7 **4.5**

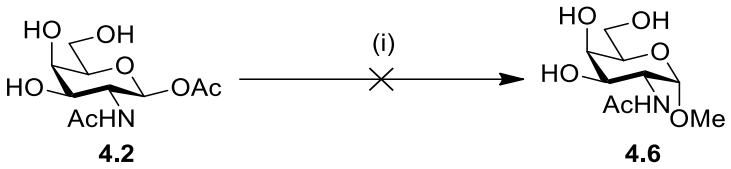
Scheme 50 - Glycosylation of MeOH to Galβ1,3GalNAc-α-OAc **4.7** (core 1-α-OAc) using WT engEF (111 mU). Conditions: engEF from *Enterococcus faecalis*, 15 % MeOH (v/v), 25 mM NaOAc, pH 6.0, 37 ˚C, 16 h, 37 %.

After 16 h, the reaction was quenched by addition of 1 M NaOH. RP-HPLC analysis of the reaction mixture indicated the formation of the desired methyl glycoside **4.5** in 37 % yield. Disappointingly, the yield obtained from the glycosylation was lower than obtained from the glycosylation of MeOH by the monosaccharide, GalNAc-α-OAc **4.1**. Due to lack of substrate, this reaction could not be optimised further.

4.4.3.2. Glycosylation of MeOH using engEF mutants

As with the mutant engBL enzymes, the glycosylation potential of the engEF D725G/A/S mutants was assessed. Firstly, we wanted to determine whether the mutant enzymes could glycosylate MeOH with the monosaccharide, GalNAc-β-OAc **4.2**. We had observed earlier that WT engEF was capable of hydrolysing GalNAc-α-OAc **4.1** and, moreover, catalysed the glycosylation of MeOH. However without the nucleophilic residue, the mutant engEF enzymes cannot form the glycosyl-enzyme intermediate. The idea was therefore that in the presence of an activated substrate containing an anomeric linkage opposite to that of the natural substrate, we may expect the mutant to be capable of catalysing the glycosylation of MeOH to give the corresponding methyl glycoside **4.6** with an inverted configuration.

Therefore, GalNAc-β-OAc **4.2** was incubated with 15 % MeOH for 16 h with the respective engEF D725G/A/S mutants (Scheme 51). The reaction was also monitored hourly by t.l.c.

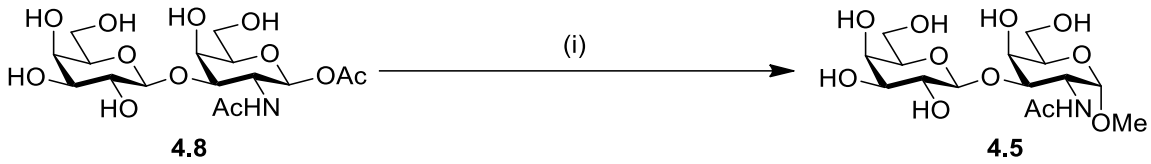


Scheme 51 - Attempted glycosylation of MeOH to GalNAc-β-OAc **4.2** using engEF D725G/A/S mutants. Conditions: engEF D725G/A/S, 15 % MeOH (v/v), 25 mM NaOAc buffer, pH 6.0, 37 ˚C, 16 h.

To our disappointment, RP-HPLC did not indicate any formation of the desired methyl glycoside **4.6** for any of the mutant reactions. T.l.c. analyses only showed the slow hydrolysis of the starting material over time. Mass spectrometry did detect the formation of the desired product **4.6**, however, since RP-HPLC and t.l.c. analyses did not show product formation, it is likely that the methyl glycoside observed was a result of direct substitution at the anomeric centre by MeOH, since product was also observed by mass spectrometry in the control reaction, where no enzyme was added.

Although these observations suggest that the mutants were unable to catalyse the transfer of MeOH to GalNAc-β-OAc **4.2**, it was still uncertain whether the enzymes were completely inactive towards activated β-glycosyl donors since it is the disaccharide, Galβ1,3GalNAc, and not the monosaccharide, that is the preferred substrate for the GH101 endo-α-*N*-acetylgalactosaminidases.

To this end, Galβ1,3GalNAc-β-OAc **4.8** (core 1-β-OAc) and 15 % MeOH were incubated with the respective engEF mutants for 16 h, at which point the reaction was quenched (Scheme 52).



4.8 **4.5**

Scheme 52 - Attempted glycosylation of MeOH to Galβ1,3GalNAc-β-OAc **4.8** using engEF D725G/A/S mutants. Conditions: engEF D725G/A/S, 15 % MeOH (v/v), 25 mM NaOAc buffer, pH 6.0, 37 ˚C, 16 h.

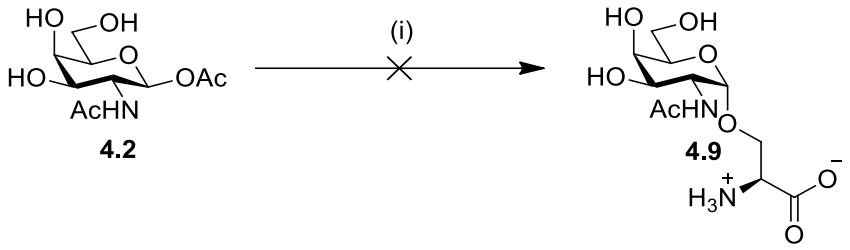
The reaction mixtures were also monitored by t.l.c. at hour intervals. However, once again no product formation was observed; in all cases RP-HPLC and mass spectrometric analyses did not indicate the presence of any of the desired methyl glycoside.

Further investigations were abandoned in light of these disappointing results.

4.4.3.3. Glycosylation of Ser-OH using engEF mutants

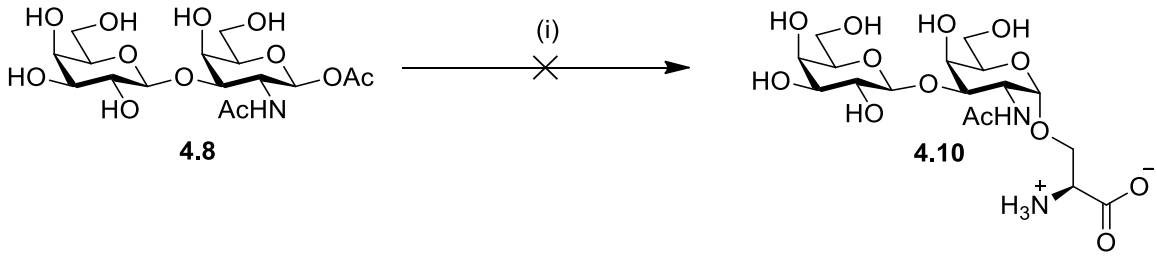
Goda *et al*[14] and Koutsioulis *et al*[25] performed extensive studies on the kinetic activity and substrate specificity of engEF from *Enterococcus faecalis.* However, to date, there are no reports detailing the use of engEF for the synthesis of *O*-linked glycosyl amino acids, glycopeptides, or glycoproteins. EngEF was selected for the work in this research due to the broad substrate specificity of the enzyme. We wanted to glycoengineer engEF to generate glycosynthases capable of glycosylating Ser/Thr residues of glycopeptides and glycoproteins, with not only the core 1-disaccharide, but also the core 3-disaccharide. However, as a preliminary step, it was appropriate to determine whether the engEF mutants could transfer activated core *O*-glycan structure onto single serine and threonine units.

Therefore, in a similar vein to the mutant engBL enzymes, engEF D725G/A/S were incubated with GalNAc-β-OAc **4.2** and 10 equivalents of Ser-OH to investigate whether the enzyme could glycosylate the unprotected amino acid with the sugar (Scheme 53).



Scheme 53 - Attempted glycosylation of Ser-OH to GalNAc-β-OAc **4.2** using engEF D725G/A/S mutants. Conditions: engEF D725G/A/S, Ser-OH (10 equiv), 25 mM NaOAc buffer, pH 6.0, 37 ˚C, 16 h.

The reactions were monitored by t.l.c. and after 16 h, the starting material was found to be completely hydrolysed with no evidence of desired product formation. The absence of the glycosyl amino acid was confirmed by RP-HPLC and mass spectrometry. To confirm the lack of glycosynthase activity was not due to substrate preference, the preferred disaccharide substrate, Galβ1,3GalNAc-β-OAc **4.8**, was also incubated with 10 equivalents of Ser-OH and the mutant enzymes (Scheme 54).



Scheme 54 - Attempted glycosylation of Ser-OH to Galβ1,3GalNAc-β-OAc **4.8** (core 1-β-OAc) using engEF D725G/A/S mutants. Conditions: engEF D725G/A/S, Ser-OH (10 equiv), 25 mM NaOAc buffer, pH 6.0, 37 ˚C, 16 h.

T.l.c. analysis of the reaction mixture at hourly intervals again did not indicate the formation of the desired glycosyl amino acid **4.10**. After 16 h, the reaction was quenched and analysed by RP-HPLC and mass spectrometry, both of which did not indicate any product formation. These results suggested that the engEF mutants were incapable of catalysing the transfer of core 1-β-OAc **4.8** to serine.

4.4.4. Conclusions

Studies by Goda *et al*[14] had reported that engEF from *Enterococcus faecalis* was capable of releasing *p*-nitrophenol from, not only Galβ1,3GalNAc-α-pNP (core 1-pNP), but also core 3-pNP demonstrating broad substrate specificity. Therefore, we were interested in using engEF for the synthesis of *O*-linked glycopeptides and glycoproteins. Chapter 2 details the expression, purification, and mutagenesis of engEF.

This chapter detailed investigations into the properties of engEF and its mutant variants. Michaelis-Menten kinetics were obtained for WT engEF. The K_M and k_{cat} values for engEF were calculated to be 53.5 μM and 76.4 s^{-1}, respectively for Galβ1,3GalNAc-α-pNP (core 1-pNP), and 42 μM and 104 s^{-1}, respectively for Galβ1,3GalNAc-α-OAc (core 1-α-OAc), indicating that glycosyl acetates are better donors for engEF compared to the *p*-nitrophenyl glycosides. WT engEF could also hydrolyse the monosaccharide, GalNAc-α-OAc.

The mutant enzymes, engEF D725G/A/S, were all completely inactive towards hydrolysis of the preferred core 1-pNP disaccharide, and could not be chemically

rescued by azide, bromide, or formate. These results suggested that the mutagenesis had affected an important residue in the active site of engEF.

WT engEF could catalyse the glycosylation of MeOH by core 1-pNP, core 1-OAc, and GalNAc-α-OAc, in yields of 56, 37, and 40 %, respectively.

None of the mutant enzymes could catalyse the glycosylation of MeOH or Ser-OH with GalNAc-β-OAc or core 1-OAc. The underlying reason for this disappointing result is not apparent. Future computational and crystallographic studies may shed light on the impact of the point mutant on the active site of the enzymes.

4.5. Overall conclusions and future work

Whilst the WT engBL and engEF enzymes were found to be active, their mutant variants were found to be catalytically inactive with the donors assayed. Disappointingly, the engBL and engEF mutant enzymes could not catalyse the glycosylation of MeOH or Ser-OH to their respective activated β-glycosyl donors. It is unclear as to which, as well as how many, aspects of the protein active site may have been affected, chemically and structurally, by the point mutations which led to these results.

Future work involving computational and crystallographic studies are required to determine the structural changes upon the substitution of the catalytic nucleophile for non-nucleophilic residues.

We also observed that glycosyl acetates were viable substrates for the GH101 endo-α-*N*-acetylgalactosaminidases examined in this work. Unfortunately, these substrates did

not contain a chromophore, necessitating the use of alternative time-consuming methodologies for assaying enzyme kinetics. Methodologies to access glycosyl esters containing a chromophore should facilitate kinetic analyses as well as yielding similar kinetic results.

4.6. References

(1) Mackenzie, L. F.; Wang, Q.; Warren, R. A. J.; Withers, S. G. *J. Am. Chem. Soc.* **1998**, *120* (22), 5583–5584.

(2) Wang, L. X.; Huang, W. *Curr. Opin. Chem. Biol.* **2009**, *13* (5-6), 592–600.

(3) Williams, S. J.; Withers, S. G. *Carbohydr. Res.* **2000**, *327* (1-2), 27–46.

(4) Wang, H.; Oman, T. J.; Zhang, R.; Garcia De Gonzalo, C. V.; Zhang, Q.; Van Der Donk, W. A. *J. Am. Chem. Soc.* **2014**, *136* (1), 84–87.

(5) Fiete, D.; Beranek, M.; Baenziger, J. U. *J. Biol. Chem.* **2012**, *4* (34), 1–18.

(6) Unverzagt, C.; Kajihara, Y. *Chem. Soc. Rev.* **2013**, *42* (10), 4408.

(7) Niggemann, J.; Kamerling, J. P.; Vliegenthart, J. F. G. *Bioorg. Med. Chem.* **1998**, *6* (9), 1605–1612.

(8) Koeller, K. M.; Wong, C.-H. *Chem. Rev.* **2000**, *100* (12), 4465–4494.

(9) Fujita, K.; Oura, F.; Nagamine, N.; Katayama, T.; Hiratake, J.; Sakata, K.; Kumagai, H.; Yamamoto, K. *J. Biol. Chem.* **2005**, *280* (45), 37415–37422.

(10) Malet, C.; Planas, A. *FEBS Lett.* **1998**, *440* (1-2), 208–212.

(11) Johnson, K. A.; Goody, R. S. *Biochemistry* **2011**, *50* (39), 8264–8269.

(12) Legler, G. *Adv. Carbohydr. Chem. Biochem.* **1990**, *48*, 319–384.

(13) Sinnott, M. L.; Souchard, I. J. *Biochem. J.* **1973**, *133* (1), 89–98.

(14) Goda, H. M.; Ushigusa, K.; Ito, H.; Okino, N.; Narimatsu, H.; Ito, M. *Biochem. Biophys. Res. Commun.* **2008**, *375* (4), 541–546.

(15) Ashida, H.; Ozawa, H.; Fujita, K.; Suzuki, S.; Yamamoto, K. *Glycoconj. J.* **2010**, *27* (1), 125–132.

(16) Suzuki, R.; Katayama, T.; Kitaoka, M.; Kumagai, H.; Wakagi, T.; Shoun, H.; Ashida, H.; Yamamoto, K.; Fushinobu, S. *J. Biochem.* **2009**, *146* (3), 389–398.

(17) Zinin, A. I.; Eneyskaya, E. V; Shabalin, K. A; Kulminskaya, A. A; Shishlyannikov, S. M.; Neustroev, K. N. *Carbohydr. Res.* **2002**, *337* (7), 635–642.

(18) Parsons, T. B.; Patel, M. K.; Boraston, A. B.; Vocadlo, D. J.; Fairbanks, A. J. *Org. Biomol. Chem.* **2010**, *8* (8), 1861–1869.

(19) International Union of Biochemistry (1965) *Enzyme Nomenclature, Recommendations 1964 of the International Union of Biochemistry*, Elsevier, Amsterdam.

(20) Hehre, E. J.; Brewer, C. F.; Genghof, D. S. *J. Biol. Chem.* **1979**, *254* (13), 5942–5950.

(21) Cobucci-Ponzano, B.; Conte, F.; Bedini, E.; Corsaro, M. M.; Parrilli, M.; Sulzenbacher, G.; Lipski, A.; Dal Piaz, F.; Lepore, L.; Rossi, M.; Moracci, M. *Chem. Biol.* **2009**, *16* (10), 1097–1108.

(22) Hommalai, G.; Withers, S. G.; Chuenchor, W.; Cairns, J. R. K.; Svasti, J. *Glycobiology* **2007**, *17* (7), 744–753.

(23) Cobucci-Ponzano, B.; Trincone, A.; Giordano, A.; Rossi, M.; Moracci, M. *Biochemistry* **2003**, *42* (32), 9525–9531.

(24) Pozzo, T.; Plaza, M.; Romero-García, J.; Faijes, M.; Karlsson, E. N.; Planas, A. *J. Mol. Catal. B Enzym.* **2014**, *107*, 132–139.

(25) Koutsioulis, D.; Landry, D.; Guthrie, E. P. *Glycobiology* **2008**, *18* (10), 799–805.

Chapter 5 – One-Pot Synthesis of Glycosides in Water

5.1. Introduction

Protecting groups are typically used in organic syntheses to 'mask' competing reactive centres around a molecule. However, this approach suffers from the need for additional synthetic steps, loss of material, and generates a huge amount of waste – not only chemical waste, but also in costs and time. Thus, there is an increasing need for syntheses that do not incorporate protecting group methodologies whilst furnishing high yields and stereoselectivities.

There are multiple examples of protecting-group-free modifications to the anomeric sugars, most of which Hanessian and Lou have covered in an excellent review article.[1] In 2009, Shoda and co-workers reported the use of the reagent 2-chloro-1,3-dimethylimidazolinium chloride (DMC) for the direct synthesis of glycosyl oxazolines from reducing sugars in water.[2] This reaction drew the attention of the glycoscience community, as glycoside formation in water is generally difficult due to competitive hydrolysis in such reactions.

The proposed mechanism[2] of oxazoline ring formation involves preferential attack of the hemiacetal hydroxyl group to DMC to yield reactive intermediate **1** (Figure 46) with a β-configuration. Intramolecular attack of the 2-acetamido group, as well as abstraction of the oxazolinium ion proton by a suitable base such as triethylamine, affords the oxazoline. The α-imidazolinium intermediate **2** can also form if the α-hemiacetal hydroxyl group attacks DMC. However, this is then hydrolysed to

regenerate the free sugar since intramolecular ring formation of the α-adduct is not possible by an S_N2 substitution pathway.

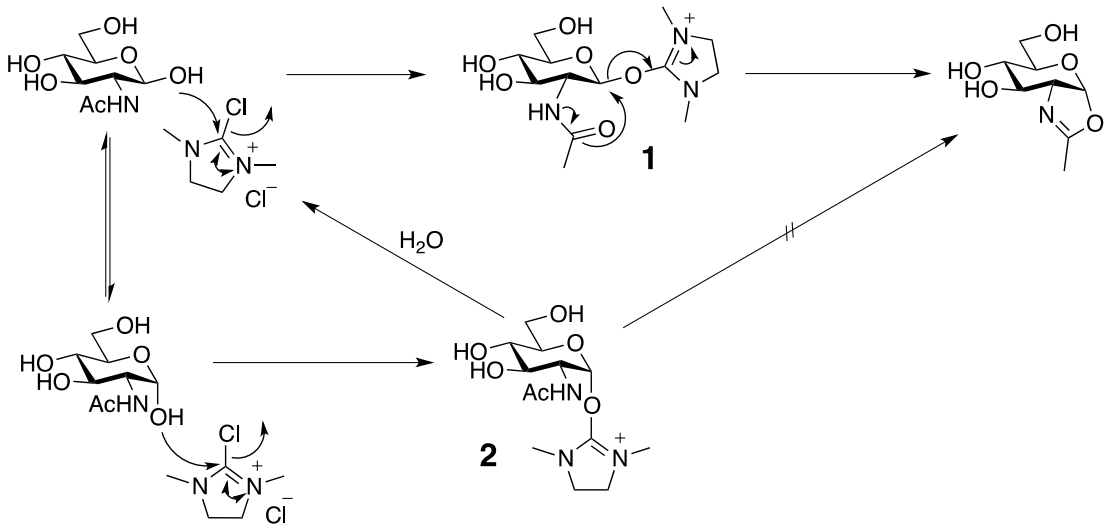


Figure 46 - Plausible mechanism for sugar oxazoline formation using DMC

Subsequently to this initial report, DMC and its derivatives have been applied to the synthesis of glycosyl azides,[3] pyridyl thioglycosides,[4] peptides linked to carbohydrates through a sulphur atom,[5] and sugar nucleoside diphosphates,[6] by intercepting the α-imidazolinium intermediate **2**[5] with different nucleophiles. This chapter outlines further developments of the DMC procedure wherein reducing sugars are activated in water in the presence of various nucleophiles to generate glycosides.

5.2. Attempted Synthesis of Various Glycosides in Water

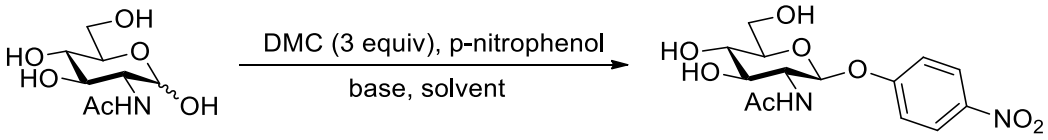
5.2.1. Nitrophenyl glycosides

The synthesis of phenyl glycosides is well reported in the literature. However, accessing these compounds typically involves several protecting group manipulations, then glycosylation of a sugar with the phenol of interest, and a final global deprotection step.[7,8]

Nitrophenyl glycosides are of importance as they are substrates for glycosidases. *p*-Nitrophenyl (pNP) and dinitrophenyl (DNP) glycosides have proven to be pivotal for monitoring and determining the kinetics of glycosidases by way of their inherent UV activity.[9,10] The synthesis of activated phenyl glycosides is also pertinent to other studies reported in this research. This section details investigations into the possible synthesis of nitrophenol glycosides in water using the aforementioned DMC methodology.

Table 14 details the conditions used in attempts to synthesise *p*-nitrophenyl glycosides of *N*-acetylglucosamine in water using DMC as an activator.

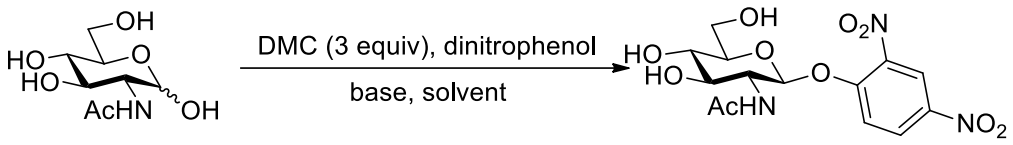
Firstly, a screen of different bases was performed[3,11] employing TEA, DIPEA, 2,6-lutidine, and K_2CO_3 (Entries 1-4). However, *p*-nitrophenol was insoluble under these conditions. A mixed solvent consisting of D_2O/MeCN (4:1) was then investigated to solubilize *p*-nitrophenol (Entries 5-10). Additionally *p*-nitrophenol was also converted to a phenolate prior to the addition of DMC (Entry 9), and the number of equivalents of *p*-nitrophenol that was added was also increased (Entry 10). However these reactions either yielded the undesired oxazoline or did not produce any reaction.



Entry	pNP (equiv)	Solvent	Base (equiv)	Result
1	3	D_2O	TEA (8)	Oxazoline only
2	3	D_2O	DIPEA (8)	Oxazoline only
3	3	D_2O	2,6-lutidine (8)	No reaction
4	3	D_2O	K_2CO_3 (8)	No reaction
5	3	D_2O/MeCN (4:1)	TEA (8)	Oxazoline only
6	3	D_2O/MeCN (4:1)	DIPEA (8)	Oxazoline only
7	3	D_2O/MeCN (4:1)	2,6-lutidine (8)	Oxazoline and Starting material
8	3	D_2O/MeCN (4:1)	K_2CO_3 (8)	No reaction
9	3	D_2O/MeCN (4:1)	NaOH (8)	No reaction
10	8	D_2O/MeCN (4:1)	TEA (8)	Oxazoline and starting material

Table 14: Conditions screened for the attempted synthesis of pNP glycosides

Withers *et al*. have also utilised di-nitrophenyl glycosides as donors for glycosynthases.[12] These are better donors than *p*-nitrophenyl glycosides due to their better leaving group properties corresponding to their lower pK_a values. The pK_a of DNP is 4.11 compared with *p*-nitrophenol, which has a pK_a of 7.16. Table 15 lists the conditions that were tested for the synthesis of DNP glycosides of *N*-acetylglucosamine in water.



Entry	DNP (equiv)	Solvent	Base (equiv)	Conversion[a]
1	10	D_2O	TEA (10)	-
2	10	D_2O	DIPEA (10)	12%
3	10	D_2O	2,6-lutidine (10)	-
4	10	D_2O	K_2CO_3 (10)	Oxazoline only
5	10	D_2O/MeCN (4:1)	TEA (10)	Oxazoline + starting material
6	10	D_2O/MeCN (4:1)	DIPEA (10)	Oxazoline + starting material
7	10	D_2O/MeCN (4:1)	2,6-lutidine (10)	Oxazoline + starting material
8	10	D_2O/MeCN (4:1)	K_2CO_3 (10)	Oxazoline + starting material
9	1	D_2O	TEA (10)	12%
10	2	D_2O	TEA (10)	7%
11	5	D_2O	TEA (10)	3%
12	20	D_2O	TEA (10)	2%

Table 15: Conditions screened for DNP glycoside synthesis
[1] Determined by ^1H NMR by comparing integrals of the anomeric proton of the products and that of the unprotected sugar in D_2O.

As previously, a screen of organic bases typically used in DMC-associated reactions was undertaken (Entries 1-4). As with *p*-nitrophenol, DNP was also found to be only

partially soluble under these conditions. No reaction occurred when TEA and 2,6-lutidine (Entries 1 & 3) were used as bases. However, using DIPEA (Entry 2) furnished the desired β-DNP glycoside in 12 % conversion (as observed by ^{1}H NMR). However the reaction did not proceed further. The use of K_2CO_3 as the base (Entry 4) completely converted GlcNAc to the undesired oxazoline. A D_2O/MeCN (4:1) solvent mixture was then used which completely solubilised the DNP. However the reactions were unsuccessful (Entries 5-8), producing only mixtures of oxazoline and recovered starting material. Increasing the DNP concentration (Entries 9-12) did result in the production of some of the desired product, but only in low conversion. Further investigations into this transformation were abandoned.

5.2.2. β-*o*-Carboxyphenyl glycosides

In 1995, Wang and Withers tested these glycosides as donors substrates for glycosynthases.[13] This type of glycoside could undergo substrate-assisted catalysis when used as a substrate for a glycosynthase derived from an inverting glycosidase (Figure 47).

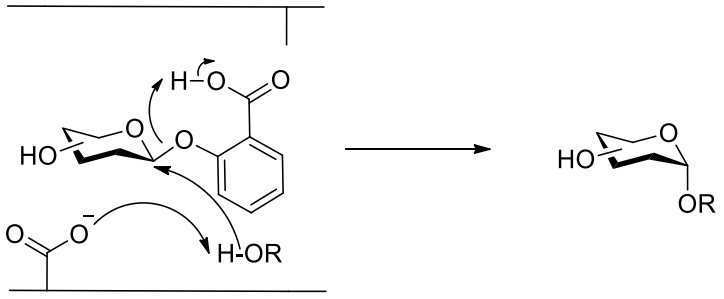
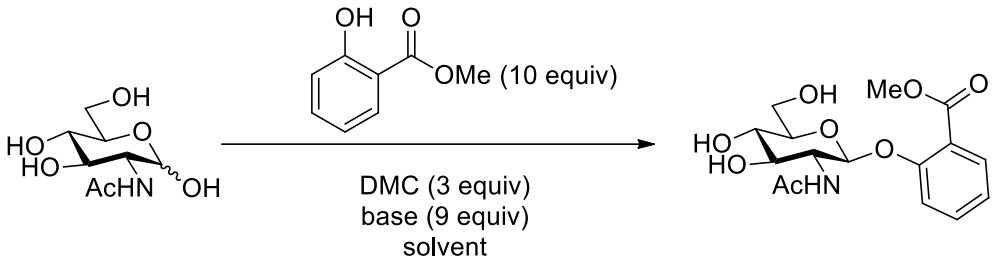


Figure 47 - Mechanism for substrate-assisted catalysis

The carboxyl group of β-*o*-carboxyphenyl glycosyl donors can act as a mimic of the general acid residue in the wild type enzyme, but which has been removed from the glycosynthase enzyme active site by mutagenesis. The acceptor nucleophilically reacts with the anomeric centre in the active site of the enzyme, which then causes the leaving phenolate to intramolecularly abstract a proton from the carboxylic acid to give the desired glycosylation product that is then released from the active site.

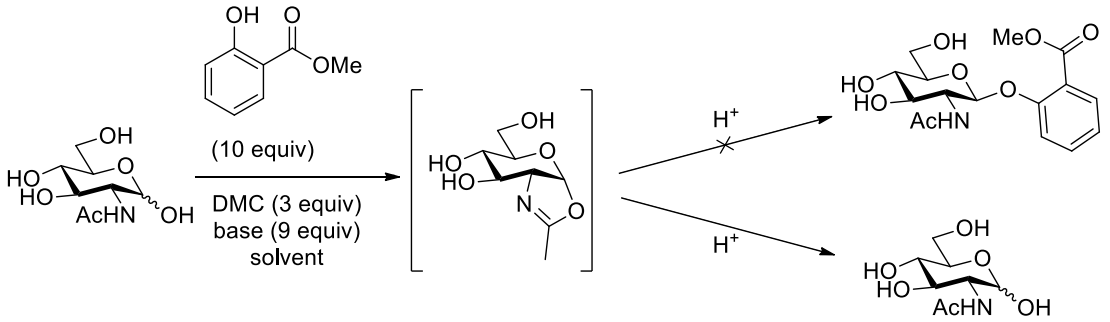
Table 16 details the set of conditions that were investigated using methyl salicylate as a nucleophile in an attempt to directly synthesise the desired β-*o*-carboxyphenyl glycosides.



Entry	Solvent	Base (+ additional reagents)	Product?[1]
1	D_2O	Triethylamine	Yes (<1)
2	D_2O	2,6-Lutidine	No
3	D_2O	K_2CO_3	No
4	D_2O	-	No
5	D_2O/acetone (1:1)	Triethylamine	No
6	D_2O/MeCN (2:1)	Triethylamine	No
7	D_2O/DMF (1:4)	Triethylamine	No
8	DMF	Triethylamine	No
9	DMSO	Triethylamine	No
10	D_2O	Triethylamine (+ LiBr)[2]	No
11	D_2O	Triethylamine (+ KI)[2]	No
12	D_2O/MeCN (4:1)	Triethylamine (+ LiBr)[2]	Yes (N.D)
13	D_2O/MeCN (4:1)	Triethylamine (+KI)[2]	Yes (5)
14	D_2O/MeCN (1:1)	Triethylamine (+ LiBr)[2]	No
15	D_2O/MeCN (1:1)	Triethylamine (+KI)[2]	No

Table 16: Glycosylation of methyl salicylate using DMC in water. N.D. Not Detected.
[1] Determined by ^1H NMR by comparing integrals of the anomeric proton of the products and that of the unprotected sugar in D_2O. Numbers in parentheses refer to the conversion yield of the product.
[2] Both 20 mol% and 100 mol% of added LiBr and KI were screened; both gave the same product conversion.

Initial attempts were made to directly glycosylate *N*-acetyl-D-glucosamine with methyl salicylate. Firstly, the reaction was performed in water, either with various added bases (Table 16, Entries 1-3) and without base (Entry 4), but in all these cases, either little or no desired product was observed. Instead, competing oxazoline formation and hydrolysis and reactions predominated. Aqueous solvent mixtures (Entries 5-7) and polar aprotic solvents (Entries 8-9) were also screened to try and reduce or eliminate the hydrolytic reactions. However these reactions was sluggish and no desired product observed. The following attempts used LiBr and KI to encourage *in situ* oxazoline ring opening (Entries 10-15), with only two of the six reactions (Entries 12-13) having low (< 5%) conversions.



Entry	Solvent	Base	Acid	Outcome
1	D_2O/MeCN (4:1)	triethylamine	1 M aq. HCl	Complete hydrolysis
2	D_2O/MeCN (6:1)	triethylamine	1 M aq. HCl	Complete hydrolysis
3	D_2O	triethylamine	*p*-TSA	Partial hydrolysis
4	D_2O	triethylamine	CSA	Partial hydrolysis

Table 17: Oxazoline opening by acid treatment in the presence of methyl salicylate

Treatment of the crude reaction mixture directly with various acids in order to try and induce acid catalysed opening of the oxazoline by methylsalicylate was also attempted, with no success (Table 17).

In light of the results observed for the attempted pNP and DNP glycoside formation using DMC in water, it was concluded that methyl salicylate was simply too poor a nucleophile in water to warrant any further investigation at this point in time.

5.2.3. Glycosyl fluorides

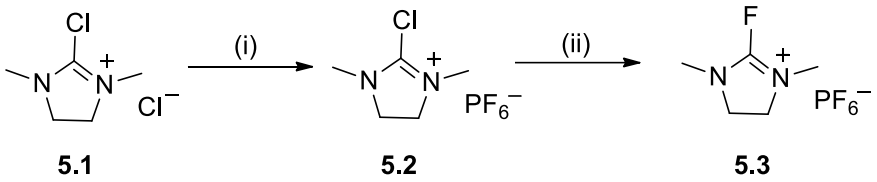
Glycosyl fluorides are known to be excellent donors in glycosylation reactions catalysed by glycosynthases.[14,15] An advantage of using a glycosyl fluoride as a donor for a glycosynthase is the small size of fluorine, which is comparable to a hydroxyl group, meaning that steric demands are minimized when using such donors.

There have been extensive efforts towards the synthesis of β-glycosyl fluorides. Notably, Helferich's original halogen exchange reaction on glycosyl bromides with silver fluoride is a reliable method that affords β-glycosyl fluorides, and is still used today.[16] Since then, the development of alternative methods to access β-glycosyl fluorides have been investigated, including the use of 2-fluoropyridinium salts, HF/pyridine complex, hexafluoropropene-amine complex, and the treatment of sugars with a free anomeric hydroxyl functionality with DAST.[17–21] A broad review has been published which details the widely used methods for the synthesis of glycosyl fluorides.[22] The primary issue that has to be overcome is the instability of β-glycosyl fluorides relative to α-glycosyl fluorides. It is known that β-glycosyl fluorides hydrolyse up to 40 times faster than α-glycosyl fluorides, which are stabilized by the

anomeric effect. This becomes a significant issue since glycosylation conditions are performed in an aqueous environment.[23] Zechel *et al*. overcame this problem by generating glycosyl fluorides *in situ* by the enzyme catalysed reaction of DNP glycosides with nucleophilic fluoride.[24]

However at the time of writing this thesis, there are still no known methods for forming glycosyl fluorides in water.

Initially, an attempt was made to synthesise a derivative of DMC that was also capable of donating fluoride. The advantage of using such a derivative is the generation of fluoride *in situ*. Therefore, it was chosen to convert DMC to 2-fluoro-1,3-dimethylimidazolium hexafluorophosphate **5.3** (DFIH, Scheme 55).

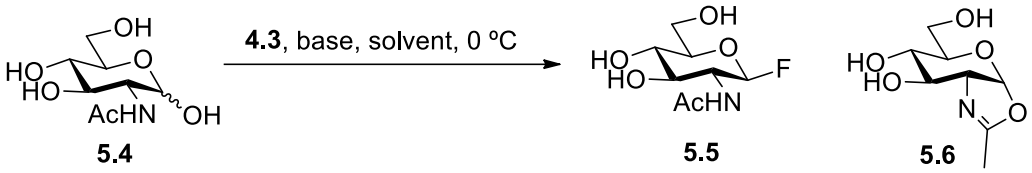


5.1 **5.2** **5.3**

Scheme 55 - Synthesis of DFIH **5.3.** Conditions: i) $NaPF_6$, MeCN, 30 mins, 99 %; ii) KF, MeCN, reflux, 4 h, 60 %.

First, DMC **5.1** was converted to the hexafluorophosphate salt **5.2** by treating with $NaPF_6$. This was then followed by displacement of chloride for fluoride by treating **5.2** with KF in MeCN and heating at reflux which gave DFIH **5.1** in 60 % yield.

N-Acetyl-D-glucosamine **5.4** was then subjected to reaction with DFIH under a variety of conditions (Table 18).



5.4 **5.5** **5.6**

Entry	Substrate (mM)	DFIH (equiv.)	Solvent	Base (equiv.)	Result
1	GlcNAc (250)	3	water	TEA (9)	Oxazoline formation only.
2	GlcNAc (250)	3	MeOH	TEA (9)	Oxazoline formation only.
3	GlcNAc (250)	3	water	-	No reaction even with heating.
4	GlcNAc (250)	3	DMSO	TEA (3)	Minor oxazoline formation. Major starting material.
5	GlcNAc (250)	3	water	DIPEA (9)	Trace amount of product but major oxazoline formation.

Table 18: Attempted fluorination reactions of GlcNAc with DFIH **5.3**

It is clear that the competing oxazoline formation **5.6** was the predominant reaction and it was quite possible that any β-fluoride formed in fact underwent immediate reaction by participation of the 2-acetamido group to give the oxazoline.

5.2.4. Conclusions

It was thought that the low nucleophilicity of fluoride, due to hydrogen bonding interactions with water, prevented glycosyl fluoride formation. It is known that hydrated fluoride ions can also activate water as a nucleophile, which would increase

the rate of hydrolysis of the imidazolium intermediate that is formed during the reaction.[25] Additionally, oxazoline ring formation was found to outcompete the desired β-fluoride formation. Due to these difficulties, further investigations into the aqueous synthesis of glycosyl fluorides was not undertaken at this point in time.

5.3. Synthesis of Glycosyl Triazoles in Water

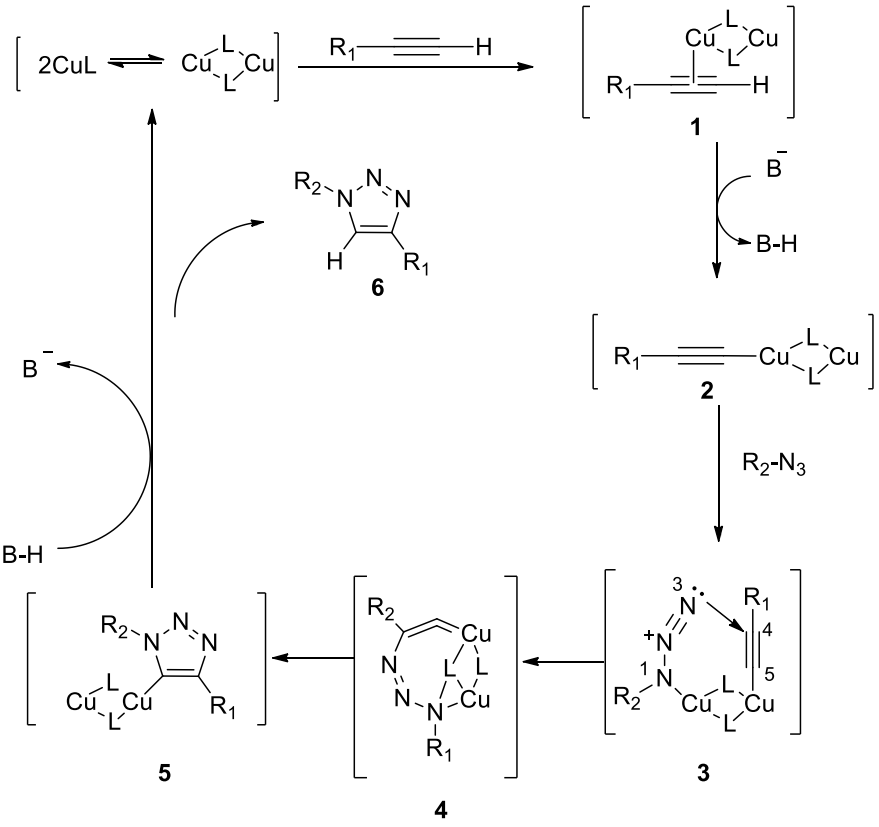
5.3.1. The Cu-catalysed Azide-alkyne Huisgen Cycloaddition (CuAAC)

The azide-alkyne Huisgen cycloaddition of organic azides and alkynes has garnered considerable attention over the past decade, especially due to the introduction of Cu(I) catalysis, which lead to a major advancement in both the regioselectivity and rate of the reaction.[26–28] In his review in 2001, Barry Sharpless defined click chemistry to be a group of reactions that "…must be modular, wide in scope, give very high yields, generate only inoffensive by-products that can be removed by non-chromatographic methods, and be stereospecific (but not necessarily enantioselective). The required process characteristics include simpler reaction conditions (ideally, the process should be insensitive to oxygen and water), readily available starting materials and reagents, the use of no solvent or a solvent that is benign (such as water) or easily removed, and simple product isolation. Purification, if required, must be by non-chromatographic methods, such as crystallization or distillation, and the product must be stable under physiological conditions".[29]

More commonly known as the Cu(I)-catalysed azide-alkyne Huisgen cycloaddition (CuAAC), the versatility of this reaction has been demonstrated by its insensitivity, robustness, and generality, as well as by a multitude of applications in the chemical and biological fields of science.[30] Reactions work under operationally simple, oxygen- and

water-tolerant conditions, and generate products in high yields with minimal requirements for product purification. Indeed, triazole chemistry has been used to modify and potentiate antitumor, antibacterial, antifungal, and antiviral compounds.[31–34] Thus, in the past decade, a multitude of catalytic conversions of acetylenes and azides in a wide range of syntheses and chemical ligations has been observed,[35–38] as well as for the synthesis of peptidomimetics and bioconjugations,[39–41] and has been touted a "green" aqueous reaction.[42] The triazole ring has been demonstrated to be effectively chemically inert to a variety of reaction conditions, such as oxidation, reduction, and hydrolysis. Furthermore its synthesis can be performed in aqueous or organic environments.[41,43]

The mechanism of the CuAAC reaction has been extensively investigated and debated.[44,45] The major energetic difference between the CuAAC and its non-catalysed counterpart is the activation energy barrier. Quantum mechanical calculations of the non-catalysed reaction show that the transition state is largely non-polarized.[46] Most of the energy is associated with the distortion energy of the azide constituting 18.1 of the 29.9 kcal/mol $\Delta G^{\dagger}$ for the reaction with acetylene. A stepwise process reduces the transition state energy by 11 kcal/mol thus achieving rate enhancement.[47] The Cu^{+} coordinates first with the acetylene π–electrons, lowering the pK_a of the acetylene proton by up to 9.8 pK_a units, allowing de-protonation to occur in an aqueous solvent without the addition of a base (Scheme 56). This coordination step is followed by an exothermic formation of a copper acetylide **2**. The Cu^{+}-acetylide complex then coordinates to the azide **3**, which is followed by rearrangement of the complex first into a metallocycle **4,** and then further into the copper-metallated triazole **5**. The Cu-triazole complex then releases the free triazole **6** and 2Cu(I)L.



Scheme 56 - Proposed catalytic cycle for the CuAAC reaction.

There have been minor revisions to this mechanism, [44,45] however the contribution of the reduction of the transition state energy to the catalysis remains unchallenged.

Although the reaction can be performed with almost any source of solvated Cu(I),[44] Cu(I) can be oxidized to Cu(II), rendering the catalyst useless. There are multiple methods to generate the active catalyst for the CuAAC reaction. One of the most common is to reduce Cu(II) salts, such as $CuSO_4 \cdot 5H_2O$, *in situ* to form Cu(I) salts. Sodium ascorbate is typically used as the reducing agent,[44] but other reducing agents, including hydrazine[48] and tris(2-carboxyethyl)phosphine (TCEP)[49] have also been used with success.

Alternatively, it is possible to directly add Cu(I) salts such as CuBr, CuI, $CuOTf.C_6H_6$, $[Cu(NCCH_3)_4][PF_6]$, to name a selection that have been used over the past few years.[50]

These methods do not require a reducing agent, but suffer from the need to be performed in a deoxygenated environment, and in an organic (or mixed) solvent system, meaning that protecting groups will probably be needed. Also, side product formation has been observed.[50,51] It can be concluded that the usage of Cu(I) salts is not as reliable as the Cu(II) procedure.

In 2005, pentamethyl cyclopentadienyl ruthenium (II) complexes (Cp*Ru), such as $Cp^*RuCl(PPh_3)_2$, were discovered as novel catalysts for click chemistry. This process, named the Ruthenium-catalysed Azide-Alkyne Cycloaddition (RuAAC), affords only 1,5-substituted 1,2,3-triazoles (Figure 48).[52]

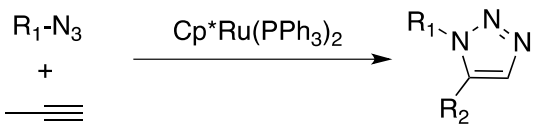


Figure 48 - Ruthenium-catalysed Azide-Alkyne Cycloaddition

Furthermore, this procedure can be used to click both terminal and internal alkynes alike. [53]

5.3.2. Biological Applications of 1,2,3-triazoles

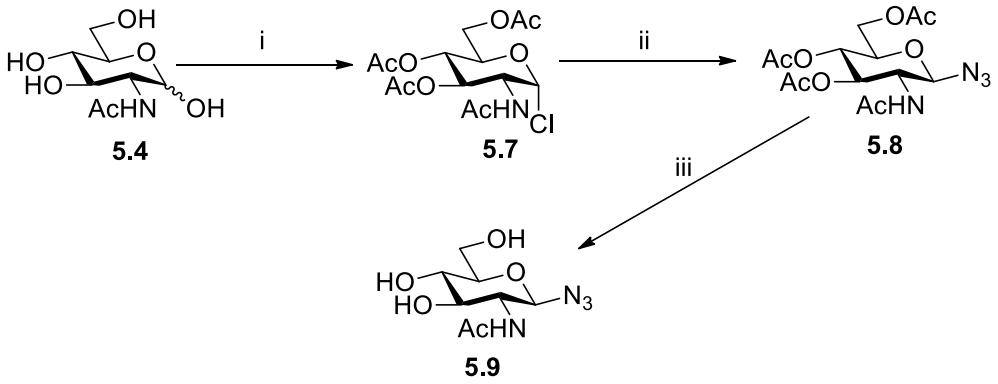
Cu(I)-catalysed triazole formation has been utilized extensively for the modification of natural products and pharmaceuticals. This methodology may allow the attachment of property modifying chemical entities to a drug in order to enhance its effects (e.g. by increasing its solubility and/or bioavailability).[54,55] The bio-isosteric similarities between triazoles and the amide bond[56] have also stimulated investigations into pharmaceutical applications of click chemistry by the generation of stable drugs and

peptidomimetics. There are several reviews available which highlight the advantages of using click chemistry for pharmaceutical applications.[57]

Carbohydrates are an important class of signalling molecules in nature, and have also been subjected to CuAAC chemistry. In fact, a prevalent method for linking sugars to other organic molecules has been to introduce an azide in the 1-, 2-, or 6-positions of the sugar and then to react this with a variety of organic alkynes.[58–60] Triazoles have also been used to link a carbohydrate functionality to different carbohydrates and amino acids.[61,62] There have also been several investigations into the application of triazoles as inhibitors of glycosidases.[63] The biological interest in triazoles stems from the fact that they are fairly easy to install and are extremely stable under standard conditions.[64,65] Their ability to tolerate oxygen, water, common conditions used for organic synthesis, biological molecules, a large range of solvent and pH's, and the reaction conditions of living systems (reducing environment, hydrolysis, etc.), makes them ideal candidates as isosteres of the amide bond in biological compounds.[44,49,64]

5.3.3. Synthesis of glycosyl azides

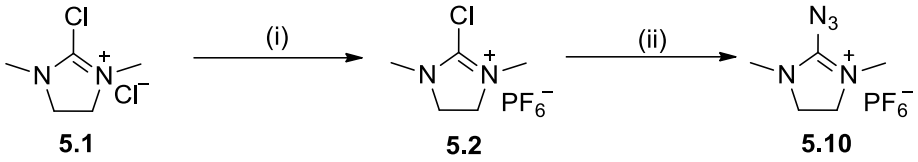
The traditional method for glycosyl azide formation involves at least three or four protecting group manipulations to furnish the desired deprotected glycosyl azide. A synthetic route typically involves converting the desired sugar, for example *N*-acetyl-D-glucosamine, into a protected glycosyl halide **5.7**, followed by a nucleophilic displacement by an azide nucleophile to give the glycosyl azide **5.8** (Scheme 57). Finally, the ester protecting groups are removed using Zemplén deacetylation to give the deprotected sugar **5.9**.



Scheme 57 - Synthesis of de-protected glycosyl azides. Conditions: i) AcCl; ii) NaN_3, sat. aq. $NaHCO_3$, $^tBuNH_4.HSO_4$, DCM; iii) Na, MeOH.

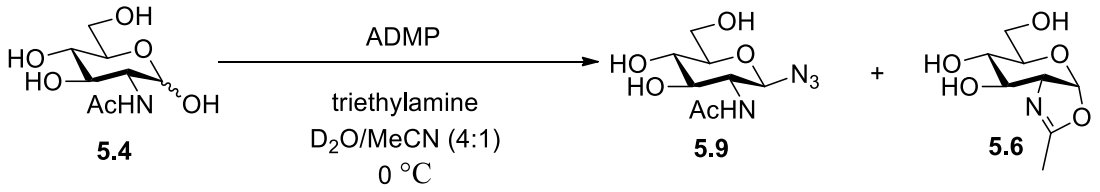
5.3.4. Synthesis of glycosyl azides using ADMP

The original Shoda method potentially generates up to two equivalents of HCl per equivalent of DMC used; one equivalent produced from deprotonation of the anomeric hydroxyl group, and one from water in the case of a hydrolysis reaction. This necessitates the use of excessive amounts of triethylamine. We proposed that if we replaced the chloride on the dimethylimidazolidinium ring with azide, it would allow for the activation of reducing sugars, as well as furnishing an azide source without the use of sodium azide. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (ADMP) has been used on several occasions in the literature as an efficient diazo-transfer reagent.[66,67] Impact sensitivity tests and friction sensitivity tests have demonstrated that ADMP is not explosive,[67] making it in fact safer to handle than sodium azide. Additionally unlike its DMC counterpart, which suffers from hygroscopicity,[67,68] ADMP is isolated as a stable crystalline solid and so it is easily handled. ADMP was synthesized following the procedure reported by Kitamura *et al.* (Scheme 58).[66] Firstly, DMC **5.1** was converted to the hexafluorophosphate salt **5.2**, and then treated with sodium azide to afford ADMP **5.10**.



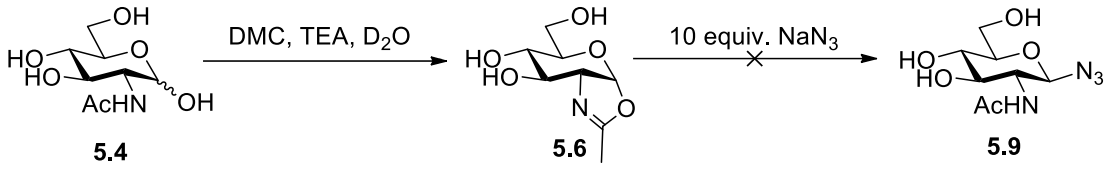
Scheme 58 - Synthesis of ADMP **5.10**. Conditions: i) $NaPF_6$, MeCN, rt, 30 mins, 99 %; ii) NaN_3, MeCN, 0 °C, 3 h, quant.

The ability of ADMP to convert sugars into glycosyl azides was tested. *N*-Acetyl-D-glucosamine **5.4** was chosen as the first substrate. Five equivalents of triethylamine was used based on the need to deprotonate the anomeric hydroxyl group as well as accounting for any hydrolysis of ADMP. Thus, *N*-acetyl-D-glucosamine was reacted with ADMP **5.10** (Scheme 59). A solvent mixture of D_2O/MeCN (4:1) was used to improve the solubility of ADMP.



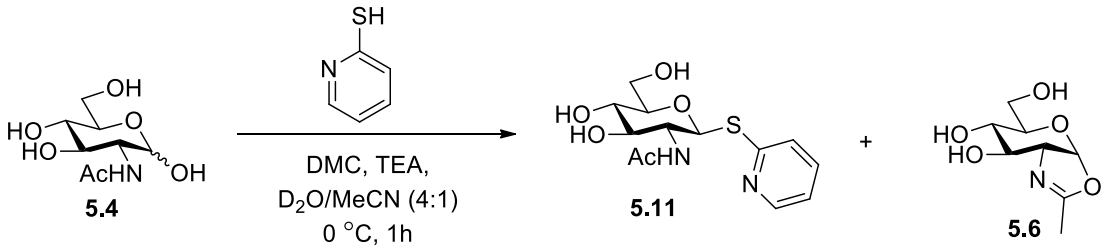
Scheme 59 - Reaction of GlcNAc **5.4** with ADMP **5.10**.

By ^{1}H NMR analysis, it was clear that competing oxazoline formation **5.6** was occurring concurrently to β-azide **5.9** formation. It was also found that oxazoline **5.6** was not converted to azide **5.9** under the reaction conditions (Scheme 60). This observation suggests that the oxazoline is not an intermediate on the pathway from the reducing sugar to the glycosyl azide, and that nucleophilic azide must either intercept either an α-chloride or α-imidazolinium intermediate to afford the glycosyl azide.



Scheme 60 - Evidence that the glycosyl oxazoline is not an intermediate on the pathway from the reducing sugar to the glycosyl azide.

Yoshida *et al*. encountered a similar problem in their attempted synthesis of pyridylthioglycosides.[4] Under similar conditions, GlcNAc was treated with 2-mercaptopyridine and DMC in the presence of triethylamine to afford the corresponding 2-pyridyl 1-thio-*N*-acetyl-D-glucosaminide **5.11** (Scheme 61), however ^{1}H NMR analysis of the reaction mixture revealed that a considerable amount of oxazoline **5.6** was formed as a by-product.



Scheme 61 - Pyridylthio-glycosidation of GlcNAc using DMC

By treating the oxazoline with 1 M HCl, they found that the resulting oxazolinium ion could be easily attacked by 2-mercaptopyridine to afford the desired β-glycoside **5.11** in high yield. In a similar vein, we found by ^{1}H NMR that when the crude reaction mixture obtained by treatment of GlcNAc with ADMP was treated dropwise with 1 M aqueous HCl until the solution was acidic, and then immediately neutralized using sat. $NaHCO_3$ solution, the anomeric oxazoline proton completely disappeared (5.9 ppm), and the desired azide was the sole product (Figure 49)

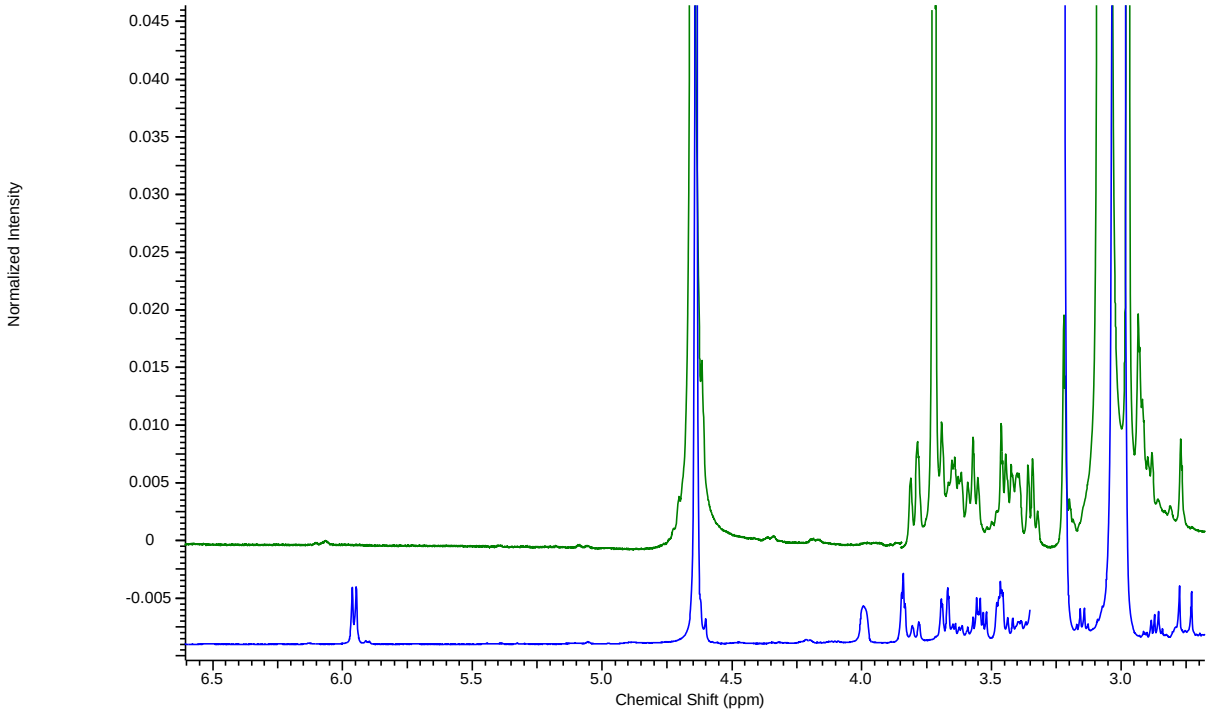


Figure 49 - ^{1}H NMR spectra of before (bottom) and after (top) acid workup of reaction of *N*-acetylglucosamine with ADMP.

5.3.5. Synthesis of glycosyl triazoles in a one-pot reaction

The next step of the investigation was to develop a one-pot click reaction. Propargyl alcohol was chosen to the alkyne coupling partner in order to optimize the reaction conditions. After the formation of the glycosyl azide was complete, propargyl alcohol, copper sulphate, and sodium ascorbate were added, and the mixture was heated to 50 °C, which led to the formation of the glycosyl triazole in excellent yields and stereoselectivities. This procedure was applied to a number of sugars ranging from mono- to trisaccharides (Table 19).

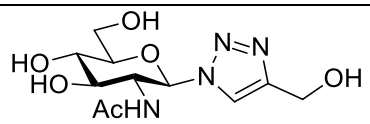
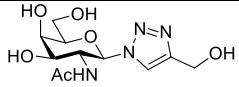
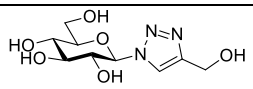
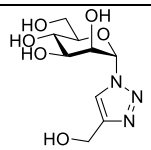
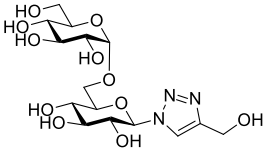
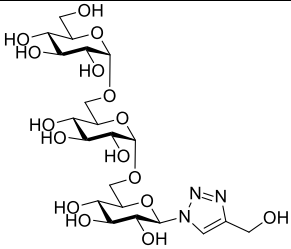
Sugar	Product	Yield	α/β
GlcNAc **5.4**	 **5.17**	86 %	β only
GalNAc **5.13**	 **5.14**	75 %	β only
Glucose **5.15**	 **5.16**	73 %	β only
Mannose **5.17**	 **5.18**	73 %	α only
Isomaltose **5.19**	 **5.20**	86 %	β only
Isomaltotriose **5.21**	 **5.22**	88 %	β only

Table 19: One-pot synthesis of glycosyl triazoles directly from corresponding reducing sugars in water. Conditions: sugar, Et_3N (5 equiv), ADMP (3 equiv) in D_2O/MeCN (4:1, 270 mM) at 0 ˚C. After 3 h, alkyne (2 equiv), $CuSO_4 \cdot 5H_2O$ (1.5 mol%) and L-ascorbic acid (0.2 equiv) were added and the mixture heated to 50 ˚C for 14 h.

The generality of the one-pot process with respect to the alkyne towards the reaction conditions was then assayed. Table 20 shows the range of alkynes tested. Entries 1-3 demonstrate that the process is extendable toward a wide variety of substituted alkynes such as in the case of a tertiary alcohol (Entry 1), carboxylic acid (Entry 2), and toward more interesting and biologically relevant applications (Entry 4). These reactions furnished the product in high yield and with complete stereoselectivity. As expected, internal alkynes are unable to undergo the cycloaddition (Entry 5) due to the inability to form the Cu-acetylide complex (absence of alkyne proton) rendering this reaction incapable of forming 1,4-disubstituted triazoles.

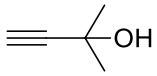
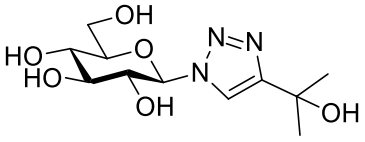
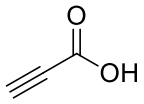
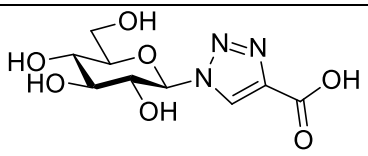
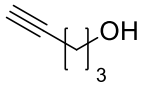
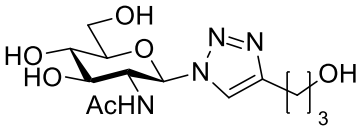
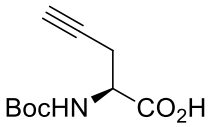
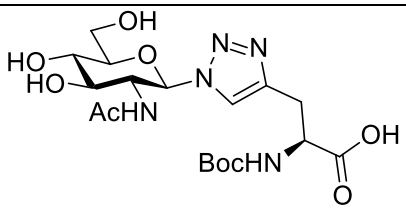
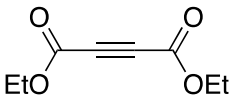
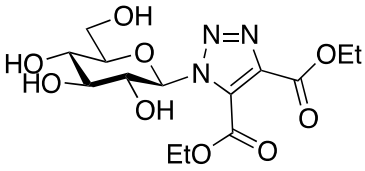
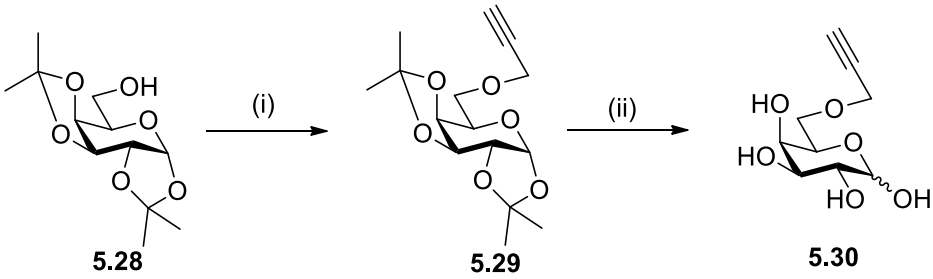
Entry	Sugar	Alkyne	Product	Yield
1	glucose **5.15**		 **5.23**	81 %
2	glucose **5.15**		 **5.24**	97 %
3	GlcNAc **5.4**		 **5.25**	95 %
4	GlcNAc **5.4**		 **5.26**	78 %
5	glucose **5.15**		 **5.27**	No reaction

Table 20: Scope of Alkyne Substrate. Conditions: sugar, Et_3N (5 equiv), ADMP (3 equiv) in D_2O/MeCN (4:1, 270 mM) at 0 ˚C. After 3 h, alkyne (2 equiv), $CuSO_4{\cdot}5H_2O$ (1.5 mol %) and L-ascorbic acid (0.2 equiv) were added and the mixture heated to 50 ˚C for 14 h.

The direct linking of carbohydrates to generate triazole-linked oligosaccharides as glycomimetics has been of interest over the past few decades.[69,70] These methods have often required multi-step synthetic strategies that are plagued by low overall yields. The

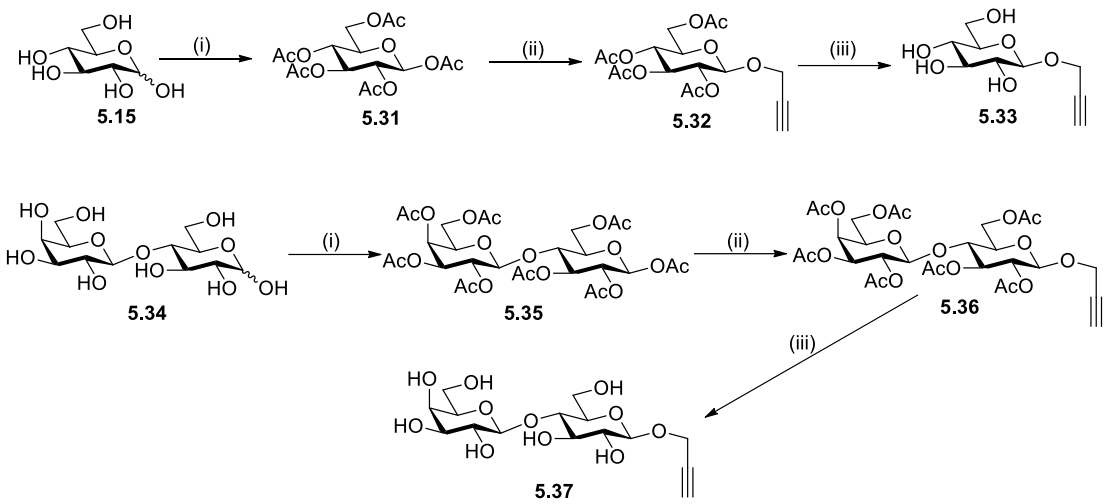
application of a one-pot method, without the need for protecting group manipulations, would therefore provide a rapid and efficient technique to assemble mimics of oligosaccharide structures in which the carbohydrate units were linked by triazoles. Therefore, several sugars incorporating an alkyne functionality were synthesized.

An alkyne functionality was easily installed at the 6-position of galactose simply by treating diacetone galactose **5.28** with propargyl bromide in the presence of K_2CO_3 to give propargylated glycoside **5.29** (Scheme 62).[71] The acetonides were then removed under acidic conditions to give alkyne **5.30**.



Scheme 62 - Synthesis of alkyne **5.30**. Conditions: (i) Propargyl bromide, KOH, MeCN, rt, 16 h, 76 %; (ii) Dowex® 50WX8 (H^+), water, 80 °C, overnight, 88 %.

Alkynes were installed linked to the anomeric centres of glucose and lactose using reported procedures by kinetic peracetylation to give glucoside **5.31** and lactoside **5.25**. These were glycosylated with propargyl in the presence of $BF_3{\cdot}Et_2O$ to give propargyl glucoside **5.32** and propargyl lactoside **5.36**. Zemplén deacetylation of **5.32** and **5.36** gave the desired propargyl glucoside **5.33** and lactoside **5.37** (Scheme 63).



Scheme 63 - Synthesis of alkyne **5.33** and **5.37**. Conditions: (i) Ac_2O, NaOAc, 120 °C, 78 % (**5.31**), 68% (**5.45**); (ii) propargyl alcohol, $BF_3{\cdot}OEt_2$, DCM, rt, 24 h, 79 % (**5.32**), 44 % (**5.36**); (iii) Na metal, MeOH, rt, 1 h, 96 % (**5.33**), quant. (**5.37**).

These were then subjected to the one-pot click reaction conditions (Table 21). Glucose was successfully clicked to propargyl ether to give a bis-linked triazole **5.38** in 84 % yield. Glucose **5.15** and galactose alkyne **5.30** were also conjugated to give a 1,6-disaccharide mimic **5.39** in 80 % yield. Glucose was linked to the *gluco* alkyne **5.33** and the *lactol* alkyne **5.37** in a head-to-head fashion to give glycosyl triazoles **5.40** and **5.41** in 84 % and 75 % yields, respectively. Pentasaccharide mimic **5.42** was synthesised by conjugating isomaltotriose **5.21** and lactosyl alkyne **5.37** under click conditions in 81 % yield.

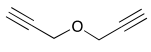
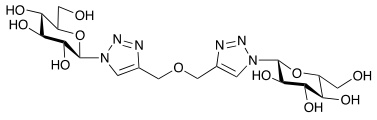
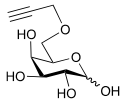
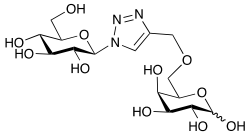
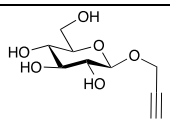
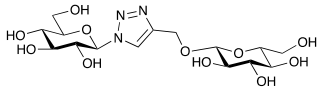
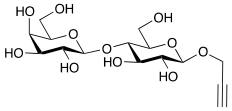
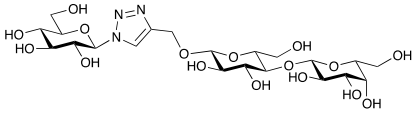
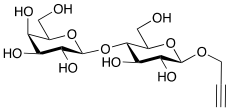
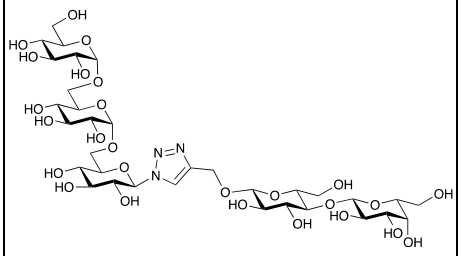
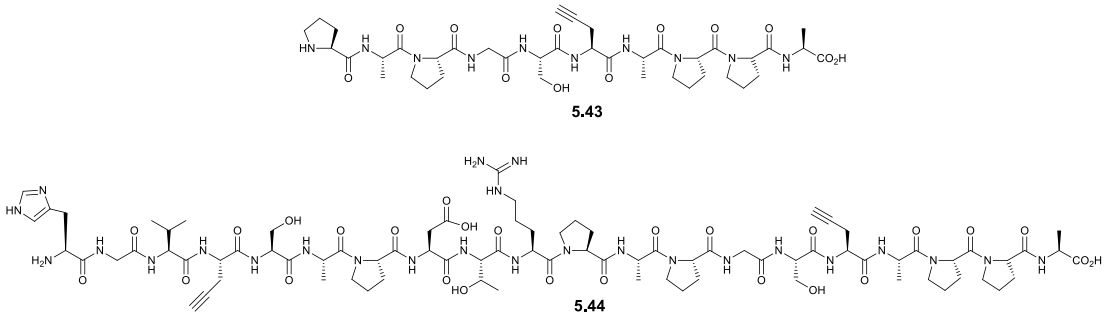
Entry	Sugar	Alkyne	Product	Yield
1	glucose **5.15**		 **5.38**	84 %
2	glucose **5.15**	 **5.30**	 **5.39**	80 %
3	glucose **5.15**	 **5.33**	 **5.40**	84 %
4	glucose **5.15**	 **5.37**	 **5.41**	75 %
5	Isomalto - triose **5.21**	 **5.37**	 **5.42**	81 %

Table 21: Direct linking of deprotected carbohydrates to access mimics of oligosaccharides. Reaction conditions: a) sugar, ADMP, Et_3N, D_2O, MeCN, 0 ˚C, 3 h, then add alkyne, $CuSO_4{\cdot}5H_2O$, L-ascorbic acid, and heat to 50 ˚C for 14 h.

Click chemistry has been applied previously to access both glycopeptides[56,72] and glycoproteins.[73] However, these reported procedures require multiple protecting group manipulations. Therefore, to further exemplify this method, an attempt to directly access glycopeptides was made. In particular, glycosylated versions of the tandem

repeat domain of the cancer-associated mucin MUC1[74,75] have shown potential as components of synthetic anticancer vaccines.[76,77]

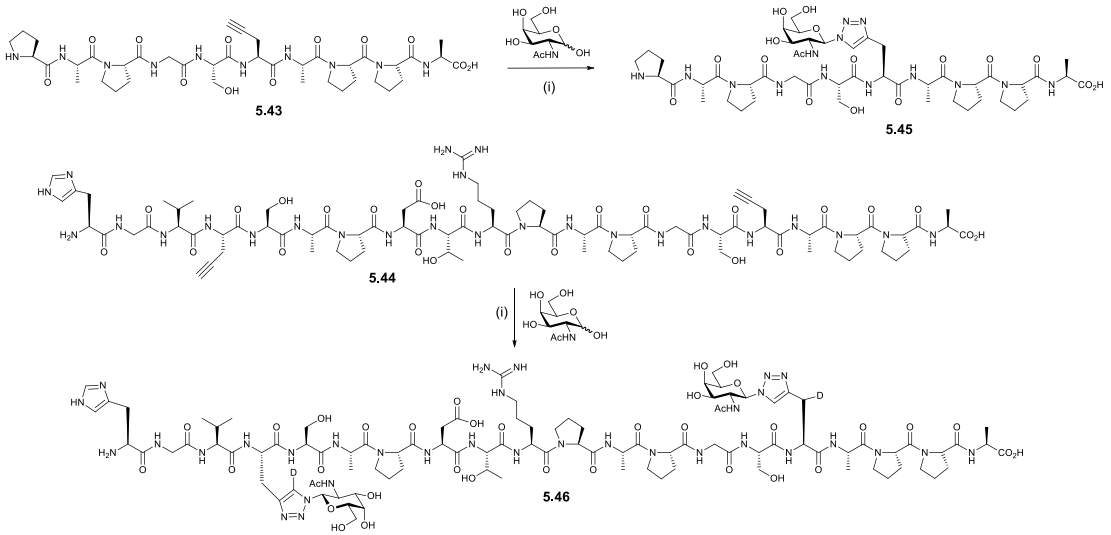
Two synthetic MUC1 peptides that incorporated propargyl glycine (Pra) were supplied to us as alkyne bearing peptides for glycoconjugation (Scheme 64).



Scheme 64 - Two MUC1 peptides, a 10-mer **5.43** and 20-mer **5.44**, incorporating Pra.

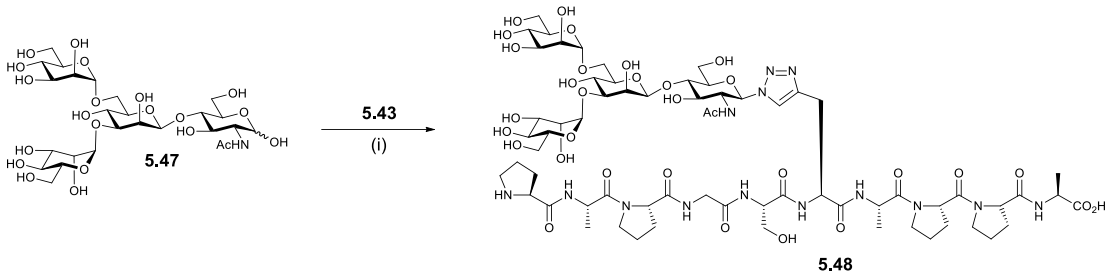
Fragment **5.43**, which corresponds to residues 11-20 in which T16 has been replaced by Pra, and the full-length tandem repeat domain **5.44**, in which both T4 and T16 have been replaced by Pra, were assembled by solid-phase peptide synthesis by the group of Brimble as previously described[78] and supplied to us.

GalNAc **5.13** was again converted to the azide using ADMP and then reacted with peptides **5.43** and **5.44**, respectively, in the presence of $CuSO_4 \cdot 5H_2O$, to yield the corresponding glycopeptides **5.45** and **5.46** in high yields (Scheme 65).



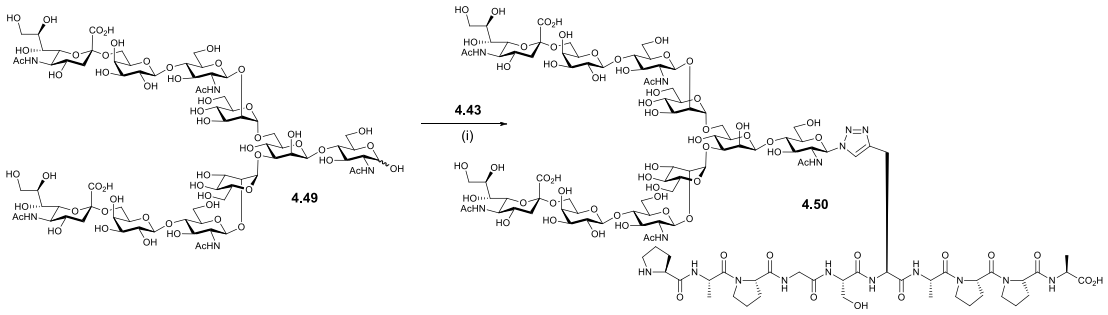
Scheme 65 - Conjugation of GalNAc with 10-mer **5.43** and 20-mer **5.44;** Reaction conditions: (i) sugar, ADMP, Et_3N, D_2O, MeCN, 0 ˚C, 3 h, then add alkyne, $CuSO_4{\cdot}5H_2O$, L-ascorbic acid, and heat to 50 ˚C for 14 h.

More complex sugars were also ligated to peptide **5.43**. The core *N*-glycan tetrasaccharide **5.47**[79] was attached to peptide **5.43** to give glycopeptide **5.48** in 42 % yield (Scheme 66)



Scheme 66 - Reaction of tetrasaccharide **5.47** with 10-mer **5.43**. Reaction conditions: i) sugar, ADMP, Et_3N, D_2O, MeCN, 0 ˚C, 3 h, then add alkyne, $CuSO_4{\cdot}5H_2O$, L-ascorbic acid, and heat to 50 ˚C for 14 h.

As a final example of the utility of this method a complex biantennary *N*-glycan **5.49** decasaccharide[80,81] was converted to the glycosyl azide using ADMP, and then reacted with peptide **5.43** in the presence of $CuSO_4.5H_2O$ to furnish glycopeptide **5.55** in 42 % yield (Scheme 67).



Scheme 67 - Ligation of decasaccharide **5.49** to 10-mer **5.43**. Reaction conditions: i) sugar, ADMP, Et_3N, D_2O, MeCN, 0 ˚C, 3 h, then add alkyne, $CuSO_4{\cdot}5H_2O$, L-ascorbic acid, and heat to 50 ˚C for 14 h.

5.3.6. Conclusion

The Cu(I)-catalysed Huisgen 1,3-dipolar cycloaddition of azides and terminal alkynes has emerged as the most popular click reaction. It has found numerous applications across a wide variety of disciplines and is of particular interest to the pharmaceutical community for its tolerance of biological conditions, most functional groups, and for the high aqueous solubility of 1,2,3-triazoles. Here, a one-pot method has been developed that allows the direct conjugation of sugars to alkynes under aqueous conditions. This reaction does not seem to have any limitations in terms of carbohydrate and alkyne coupling partner. Its simplicity has the potential to find numerable applications in the field of carbohydrate chemistry.

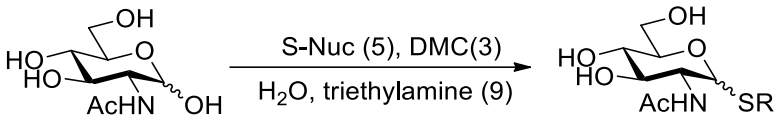
5.4. One-Pot Synthesis of Glycosyl Thiols in water

5.4.1. Introduction

Glycosyl thiols have been used previously as useful building blocks for the synthesis of glycoconjugates that may be considered to be analogues of glycopeptides and glycoproteins.[82–85] Given their potential uses, there is a drive to discover less protracted methods for the synthesis of glycosyl thiols.

5.4.2. Synthesis of Glycosyl Thioacetates in Water using DMC

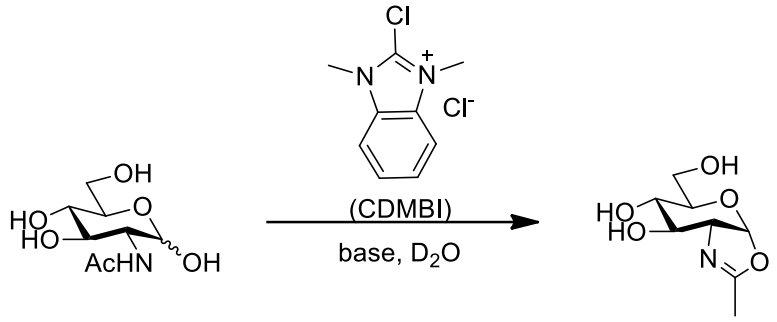
A series of sulfur nucleophiles were screened for their ability to react with GlcNAc when activated with DMC (Table 22).



Entry	S-Nuc	R	Observations[1]
1	Na_2S	H	No reaction
2	NaSH	H	No reaction
3	$SC(NH_2)_2$	$C(NH_2)_2$	No reaction
4	KSAc	Ac	No reaction
5	AcSH	Ac	New product formed

Table 22: Sulfur nucleophiles.

Initially, it was thought that introducing a thiol directly by using sodium sulfide would be the easiest option (Entry 1); unfortunately this yielded no desired product. In reported procedures, it was found that bases with high pK_a's retarded the progress of the respective desired reactions.[68] For example, when GlcNAc was reacted with 2-chloro-1,3-dimethyl-1H-benzimidazol-3-ium chloride (CDMBI) using sodium hydroxide (pK_a ~13), sodium carbonate (pK_a ~ 10.3), and potassium carbonate (pK_a ~ 10.3) as bases, the yields did not exceed 65 %, with yields decreasing to 24 % in the case of sodium hydroxide (Scheme 68).



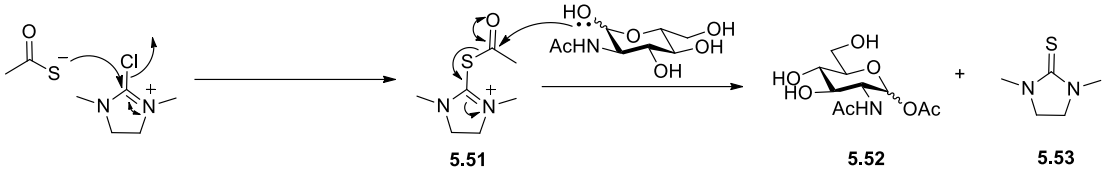
Scheme 68 - Oxazoline formation using CDMBI

Therefore, the lack of product formation can be attributed to the high pK_a of sulfides ($pK_a > 14$). Sodium hydrogen sulfide was also investigated (Entry 2), but no reaction progress could be observed by TLC and ^{1}H NMR. At this point, investigations into direct routes to the glycosyl thiol were abandoned.

Alternative sulfur nucleophiles were investigated, where the thioglycoside that would be generated could then be de-protected to furnish the desired glycosyl thiol. Thiourea[86] (Entry 3) and potassium thioacetate[87] (Entry 4) have been used previously to introduce thiols at the anomeric centre. However they proved to be unsuccessful for this reaction.

Finally, thioacetic acid was used as the source of sulfur (Entry 5). GlcNAc, triethylamine, and thioacetic acid were first dissolved in D_2O and cooled to 0 °C. DMC was then added portionwise to the reaction mixture. After 30 mins, TLC analysis indicated the formation of a major product and ^{1}H NMR analysis indicated the expected downfield shift of the anomeric proton from 4.6 ppm and 5.1 ppm to 5.5 ppm and 6.1 ppm for H-1β and H-1α, respectively. However, mass spectrometric analysis of the isolated material indicated that the mass of the product was 286.0904, which corresponded to that mass of anomeric *O*-acetate. Moreover, the product was then treated with 1 M NaOH, and after isolation, was found to be *N*-acetyl-D-glucosamine.

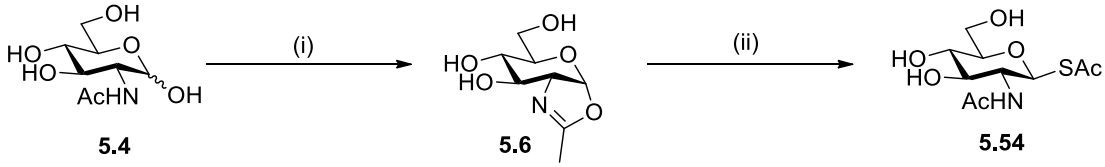
A mechanistic rationale for the unexpected generation of the glycosyl acetate is shown in Scheme 69.



Scheme 69 - Mechanism of glycosyl acetate formation using DMC.

It appeared that the thiolate reacted with DMC directly to generate an activated thioester intermediate **5.51**. Following this, the anomeric hydroxyl group of GlcNAc reacted directly with the thioester to generate the glycosyl acetate **5.52** and the thiourea **5.53** as the by-product. Further developments of this unexpected reaction will be discussed in Section 5.6.

It was then thought that changing the order of addition may help bypass the problem of forming the glycosyl acetate. The above reaction was therefore repeated but the order of adding the thioacetic acid was changed - in this second case, GlcNAc and triethylamine were stirred in H_2O and cooled to 0 °C, followed by the addition of DMC (Scheme 70).



Scheme 70 - Synthesis of thioglycoside **5.54**. Reaction conditions: (i) TEA, H_2O, 0 °C, then DMC, 30 mins; (ii) AcSH (15 equiv), 90 %.

As anticipated, the glycosyl oxazoline **5.6** was formed as the major product. The crude reaction mixture was then treated directly with thioacetic acid. As was the case for the oxazoline opening in the synthesis of glycosyl azides, the oxazoline was opened by S_N2 attack of an external nucleophile, in this case thiolate. The desired anomeric thioacetate **5,54** was formed exclusively as the β-anomer.

Further work on the generation of glycosyl thiols based on these initial findings is currently being undertaken by other members of the Fairbanks group.

5.4.3. Conclusion

A method for the one-pot direct aqueous synthesis of a glycosyl thioacetate from un-protected GlcNAc has been discovered. Simple de-esterification under basic conditions will yield the desired glycosyl thiol, which can then be further advanced by attachment to peptides and proteins using existing methods.[88] This reaction may provide many opportunities in chemical biology, and since the direct conversion of naturally occurring sugars and to glycosyl thiols and their subsequent attachment to peptides and proteins would avoid the tedious syntheses that are generally required to generate biologically-relevant glycoconjugates.

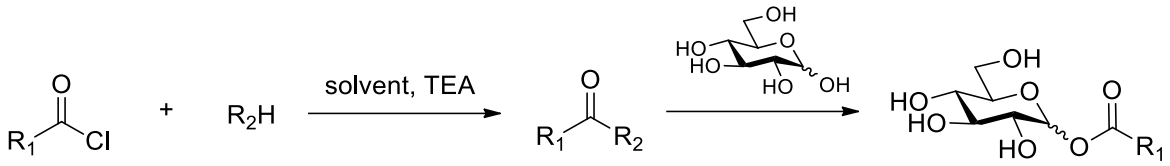
5.5. One-Pot Synthesis of Glycosyl Acetates in water

5.5.1. Introduction

The regioselective introduction of protecting groups is of central importance in carbohydrate chemistry. Acyl groups, especially acetyl and benzoyl, are widely used electron-withdrawing protecting groups. In particular, ester protection of the 2-OH of a glycosyl donor furnishes the 1,2-*trans* glycosylation product by neighbouring group

participation. There have been multiple efforts at regioselective acylation of sugars, however these methodologies have mostly focused on the selective acylation of sugar 2,3- or 4,6-diols.[89–93] Indeed most of the known derivatives of glucose in which the anomeric centre is selectively acylated have been obtained *via* acylation of 2,3,4,6-tetra-*O*-benzyl-D-glucopyranose, which occurs usually with low stereoselectivity. Alternatively, ethyl 2,3,4,6-*O*-benzyl-1-thio-β-D-galactopyranoside was reacted with *N*-iodosuccinimide and AcOH in DCM to furnished the 1-*O*-acetate in excellent yield, but as a 1:1 mixture of anomers that could not be separated.[94] Coupling anomeric mixtures of *O*-glycosyl trichloroacetimidates with carboxylic acids in non-polar solvents has also produced 1-*O*-acyl sugars in high yield, and complete inversion of stereochemistry.[95] In this latter case, the β-anomer could be separated by crystallization. Hydrogenolytic debenzylation then gave the de-protected galactoside.

Plusquellec *et al.* demonstrated an elegant method for regioselective anomeric acylation where acyl chlorides were reacted with mercaptothiazoline (Entry 1), mercaptobenzothiazole (Entry 2), *p*-nitrophenol (Entry 3), and 8-hydroxyquinoline (Entry 4) to generate the corresponding activated *S*-thioesters, amides, and esters (Figure 50).[96]



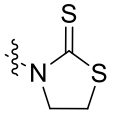
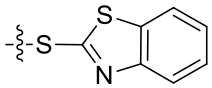
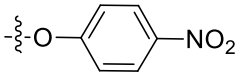
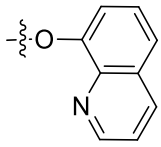
Entry	R_2
1	
2	
3	
4	

Figure 50 – Generation of reactive species for regioselective anomeric acylation by Plusquellec *et al*.[96]

This was then reacted with an unprotected sugar to give the acylated glycoside in moderate yields. However, some reactions needed to be performed at elevated temperatures (up to 70 ºC) and for long periods (up to 72 h). Moreover, the reactions were performed in an organic solvent (pyridine).

The regioselective acylation of the anomeric hydroxyl group is therefore still quite a challenge for chemists.[97] To date, there have been no published methods for accessing 1-*O*-acetyl glycosides in water. Here, I present a method for the facile access of 1-*O*-acetyl glycosides in water with perfect regiochemical control and, in some cases, stereocontrol.

5.5.2. Synthesis of Glycosyl Acetates in Water using DMC

As mentioned in Section 5.5.2, an attempt to synthesise glycosyl thiols led to the discovery of an unexpected reaction (Scheme 69). Herein stirring *N*-acetyl-D-glucosamine **5.4** with thioacetic acid and triethylamine, followed by the addition of DMC, led to selective acetylation of the anomeric centre. This reaction outcome was due to the preferential attack of the thiolate on DMC to generate a thioester intermediate **5.51**. The thioester was then attacked by the anomeric hydroxyl group, resulting in a transfer of the acetyl functionality onto the sugar and ejecting a thiourea as the by-product.

It was found that this method could be applied to various reducing sugars to generate products in which the anomeric centre had been regioselectively acetylated. Table 23 details the various free sugars that were subjected to acetylation.

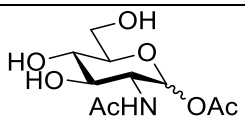
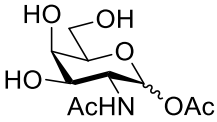
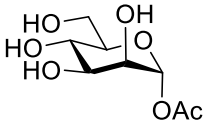
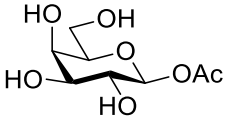
Sugar	Product	Yield	α/β
GlcNAc	 **5.52**	86 %	3:2
GalNAc	 **5.55**	86 %	1:1
Mannose	 **5.56**	73 %	α only
Galactose	 **5.58**	85 %	β only

Table 23: Direct aqueous synthesis of glycosyl acetates. Reaction conditions: sugar, water, TEA, AcSH, 0 °C, then DMC, 30 mins.

5.5.3. Conclusion

A method for the direct synthesis of otherwise unprotected glycosyl acetates from the corresponding reducing sugars in water has been achieved. Unprotected glycosyl acetates are interesting in that they have significant potential for use as activated donors for glycosidase and glycosynthases enzymes. Their use as glycosyl donors for glycosidases and glycosynthases is demonstrated in Chapter 4.

5.6. References

(1) Hanessian, S.; Lou, B. *Chem. Rev.* **2000**, *100* (12), 4443–4463.

(2) Noguchi, M.; Tanaka, T.; Gyakushi, H.; Kobayashi, A.; Shoda, S. -I. *J. Org. Chem.* **2009**, *74* (5), 2210–2212.

(3) Tanaka, T.; Nagai, H.; Noguchi, M.; Kobayashi, A.; Shoda, S.-I. *Chem. Commun.* **2009**, *23*, 3378–3379.

(4) Yoshida, N.; Noguchi, M.; Tanaka, T.; Matsumoto, T.; Aida, N.; Ishihara, M.; Kobayashi, A.; Shoda, S. -I. *Chem. Asian J.* **2011**, *6* (7), 1876–1885.

(5) Novoa, A.; Barluenga, S.; Serba, C.; Winssinger, N. *Chem. Commun.* **2013**, *49* (69), 7608–7610.

(6) Tanaka, H.; Yoshimura, Y.; Jürgensen, M. R.; Cuesta-Seijo, J. A.; Hindsgaul, O. *Angew. Chem. Int. Ed.* **2012**, *51* (46), 11531–11534.

(7) Ryzhov, I. M.; Korchagina, E. Y.; Popova, I. S.; Bovin, N. V. *Carbohydr. Res.* **2012**, *351*, 17–25.

(8) Pal, K. B.; Mukhopadhyay, B. *Carbohydr. Res.* **2014**, *400*, 9–13.

(9) Ashida, H.; Ozawa, H.; Fujita, K.; Suzuki, S.; Yamamoto, K. *Glycoconj. J.* **2010**, *27* (1), 125–132.

(10) Hollinger, M.; Abraha, F.; Oscarson, S. *Carbohydr. Res.* **2011**, *346* (12), 1454–1466.

(11) Shoda, S.; Izumi, R.; Suenaga, M.; Saito, K.; Fujita, M. *Chem. Lett.* **2002**,.*2*, 150–150.

(12) Chen, H. M.; Withers, S. G. *ChemBioChem* **2007**, *8* (7), 719–722.

(13) Wang, Q.; Withers, S. G. *J. Am. Chem. Soc.* **1995**, *117* (40), 10137–10138.

(14) Williams, S. J.; Withers, S. G. *Carbohydr. Res.* **2000**, *327* (1-2), 27–46.

(15) Williams, S. J.; Withers, S. G. *Aust. J. Chem.* **2002**, *55*, 3–12.

(16) Helferich, B.; Gootz, R. *Ber. Dtsch. Chem. Ges.* **1929**, *62* (9), 2505–2507.

(17) Mukaiyama, T.; Hashimoto, Y.; Shoda, S. *Chem. Lett.* **1983**, *6*, 935–938.

(18) Hayashi, M.; Hashimoto, S.; Noyori, R. *Chem. Lett.* **1984**, *10*, 1747–1750.

(19) Araki, Y.; Watanabe, K.; Kuan, F.-H.; Itoh, K.; Kobayashi, N.; Ishido, Y. *Carbohydr. Res.* **1984**, *127* (1), C5–C9.

(20) Posner, G. H.; Haines, S. R. *Tetrahedron Lett.* **1985**, *26* (1), 5–8.

(21) Rosenbrook, W.; Riley, D. A.; Lartey, P. A. *Tetrahedron Lett.* **1985**, *26* (1), 3–4.

(22) Yokoyama, M. *Carbohydr. Res.* **2000**, *327* (1-2), 5–14.

(23) Konstantinidis, A; Sinnott, M. L. *Biochem. J.* **1991**, *279*, 587–593.

(24) Zechel, D. L.; Reid, S. P.; Nashiru, O.; Mayer, C.; Stoll, D.; Jakeman, D. L.; Warren, R. A. J.; Withers, S. G. *J. Am. Chem. Soc.* **2001**, *123* (18), 4350–4351.

(25) Furuya, T.; Kuttruff, C. A; Ritter, T. *Curr. Opin. Drug Disc.* **2008**, *11* (6), 803–819.

(26) Huisgen, R. *Pure and Applied Chemistry*. **1989**, 613–628.

(27) Tornøe, C. W.; Christensen, C.; Meldal, M. *J. Org. Chem.* **2002**, *67* (9), 3057–3064.

(28) Huisgen, R.; Szeimies, G.; Möbius, L. *Chem. Ber.* **1967**, *100* (8), 2494–2507.

(29) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2001**, *40* (11), 2004–2021.

(30) Meldal, M.; Tomøe, C. W. *Chem. Rev.* **2008**, *108* (8), 2952–3015.

(31) Duval, R. A.; Poupon, E.; Romero, V.; Peris, E.; Lewin, G.; Cortes, D.; Brandt, U.; Hocquemiller, R. *Tetrahedron* **2006**, *62* (26), 6248–6257.

(32) Fu, X.; Albermann, C.; Zhang, C.; Thorson, J. S. *Org. Lett.* **2005**, *7* (8), 1513–1515.

(33) Li, J.; Zheng, M.; Tang, W.; He, P.-L.; Zhu, W.; Li, T.; Zuo, J.-P.; Liu, H.; Jiang, H. *Bioorg. Med. Chem. Lett.* **2006**, *16* (19), 5009–5013.

(34) Pore, V. S.; Aher, N. G.; Kumar, M.; Shukla, P. K. *Tetrahedron* **2006**, *62* (48), 11178–11186.

(35) Evans, R. A. *Aust. J. Chem.* **2007**, *60* (6), 384.

(36) Goodall, G. W.; Hayes, W. *Chem. Soc. Rev.* **2006**, *35* (3), 280–312.

(37) Hawker, C. J. *Science.* **2005**, *309* (5738), 1200–1205.

(38) Yeo, D. S. Y.; Srinivasan, R.; Chen, G. Y. J.; Yao, S. Q. *Chem. Eur. J.* **2004**, *10* (19), 4664–4672.

(39) Angell, Y. L.; Burgess, K. *Chem. Soc. Rev.* **2007**, *36* (10), 1674.

(40) Miller, N.; Williams, G. M.; Brimble, M. A. *Org. Lett.* **2009**, *11* (11), 2409–2412.

(41) Wilkinson, B. L.; Long, H.; Sim, E.; Fairbanks, A. J. *Bioorg. Med. Chem. Lett.* **2008**, *18* (23), 6265–6267.

(42) Chen, L.; Li, C.-J. *Adv. Synth. Catal.* **2006**, *348* (12-13), 1459–1484.

(43) Jiang, Y.; Kong, D.; Zhao, J.; Zhang, W.; Xu, W.; Li, W.; Xu, G. *Tetrahedron Lett.* **2014**, *55* (15), 2410–2414.

(44) Bock, V. D.; Hiemstra, H.; Van Maarseveen, J. H. *Eur. J. Org. Chem.* **2006**, *1*, 51–68.

(45) Rodionov, V. O.; Fokin, V. V.; Finn, M. G. *Angew. Chem. Int. Ed.* **2005**, *44* (15), 2210–2215.

(46) Jones, G. O.; Ess, D. H.; Houk, K. N. *Helv. Chim. Acta.* **2005**, *88* (7), 1702–1710.

(47) Himo, F.; Lovell, T.; Hilgraf, R.; Rostovtsev, V. V.; Noodleman, L.; Sharpless, K. B.; Fokin, V. V. *J. Am. Chem. Soc.* **2005**, *127* (1), 210–216.

(48) Golas, P. L.; Tsarevsky, N. V.; Sumerlin, B. S.; Matyjaszewski, K. *Macromolecules* **2006**, *39* (19), 6451–6457.

(49) Zhan, W.; Barnhill, H. N.; Sivakumar, K.; Tian, H.; Wang, Q. *Tetrahedron Lett.* **2005**, *46* (10), 1691–1695.

(50) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2002**, *41* (14), 2596–2599.

(51) Horne, W. S.; Stout, C. D.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2003**, *125* (31), 9372–9376.

(52) Zhang, L.; Chen, X.; Xue, P.; Sun, H. H. Y.; Williams, I. D.; Sharpless, K. B.; Fokin, V. V.; Jia, G. *J. Am. Chem. Soc.* **2005**, *127* (46), 15998–15999.

(53) Pirali, T.; Gatti, S.; Di Brisco, R.; Tacchi, S.; Zaninetti, R.; Brunelli, E.; Massarotti, A.; Sorba, G.; Canonico, P. L.; Moro, L.; Genazzani, A. A.; Tron, G. C.; Billington, R. A. *ChemMedChem* **2007**, *2* (4), 437–440.

(54) Aher, N. G.; Pore, V. S. *Synlett* **2005**, *14*, 2155–2158.

(55) Zinzalla, G.; Milroy, L.-G.; Ley, S. V. *Org. Biomol. Chem.* **2006**, *4* (10), 1977.

(56) Li, H.; Aneja, R.; Chaiken, I. *Molecules* **2013**, *18* (8), 9797–9817.

(57) Hein, C. D.; Liu, X.-M.; Wang, D. *Pharm. Res.* **2008**, *25* (10), 2216–2230.

(58) Wilkinson, B. L.; Bornaghi, L. F.; Houston, T. A.; Innocente, A.; Supuran, C. T.; Poulsen, S. A. *J. Med. Chem.* **2006**, *49* (22), 6539–6548.

(59) Wilkinson, B. L.; Bornaghi, L. F.; Poulsen, S.-A.; Houston, T. A. *Tetrahedron* **2006**, *62* (34), 8115–8125.

(60) Wróblewski, A. E.; Głowacka, I. E. *Tetrahedron: Asymmetry* **2005**, *16* (24), 4056–4064.

(61) Gouin, S. G.; Bultel, L.; Falentin, C.; Kovensky, J. *Eur. J. Org. Chem.* **2007**, *2007* (7), 1160–1167.

(62) Hotha, S.; Kashyap, S. *J. Org. Chem.* **2006**, *71* (1), 364–367.

(63) Dedola, S.; Hughes, D. L.; Nepogodiev, S. A.; Rejzek, M.; Field, R. A. *Carbohydr. Res.* **2010**, *345* (9), 1123–1134.

(64) Kolb, H. C.; Sharpless, K. B. *Drug Discov. Today* **2003**, *8* (24), 1128–1137.

(65) Bräse, S.; Gil, C.; Knepper, K.; Zimmermann, V. *Angew. Chem. Int. Ed.* **2005**, *44* (33), 5188–5240.

(66) Kitamura, M.; Tashiro, N.; Miyagawa, S.; Okauchi, T. *Synthesis* **2011**, *7*, 1037–1044.

(67) Kitamura, M.; Kato, S.; Yano, M.; Tashiro, N.; Shiratake, Y.; Sando, M.; Okauchi, T. *Org. Biomol. Chem.* **2014**, *12* (25), 4397–4406.

(68) Noguchi, M.; Fujieda, T.; Huang, W. C.; Ishihara, M.; Kobayashi, A.; Shoda, S. I. *Helv. Chim. Acta* **2012**, *95* (10), 1928–1936.

(69) Lee, D. J.; Mandai, K.; Harris, P. W. R.; Brimble, M. A.; Kent, S. B. H.; Mandal, K.; Harris, P. W. R.; Brimble, M. A.; Kent, S. B. H. *Org. Lett.* **2009**, *11* (22), 5270–5273.

(70) Fernández-González, M.; Boutureira, O.; Bernardes, G. J. L.; Chalker, J. M.; Young, M. A.; Errey, J. C.; Davis, B. G. *Chem. Sci.* **2010**, *1* (6), 709.

(71) Ito, H.; Kamachi, T.; Yashima, E. *Chem. Commun.* **2012**, *48* (45), 5650.

(72) Tomabechi, Y. *Trends Glycosci. Glycotechnol.* **2015**, *27* (154), 63–65.

(73) Wang, H.; Huang, W.; Orwenyo, J.; Banerjee, A.; Vasta, G. R.; Wang, L. X. *Bioorg. Med. Chem.* **2013**, *21* (7), 2037–2044.

(74) Hanisch, F.-G.; Muller, S. *Glycobiology* **2000**, *10* (5), 439–449.

(75) Sherblom, A. P.; Moody, C. E. *Cancer Res.* **1986**, *46* (9), 4543–4536.

(76) Kaiser, A.; Gaidzik, N.; Westerlind, U.; Kowalczyk, D.; Hobel, A.; Schmitt, E.; Kunz, H. *Angew. Chem. Int. Ed.* **2009**, *48* (41), 7551–7555.

(77) Lakshminarayanan, V.; Thompson, P.; Wolfert, M. A.; Buskas, T.; Bradley, J. M.; Pathangey, L. B.; Madsen, C. S.; Cohen, P. A.; Gendler, S. J.; Boons, G.-J. *Proc. Natl. Acad. Sci.* **2012**, *109* (1), 261–266.

(78) Lee, D. J.; Harris, P. W. R.; Brimble, M. A. *Org. Biomol. Chem.* **2011**, *9* (5), 1621.

(79) Rising, T. W. D. F.; Heidecke, C. D.; Moir, J. W. B.; Ling, Z.; Fairbanks, A. J. *Chem. Eur. J.* **2008**, *14* (21), 6444–6464. The tetrasaccharide was synthesised by Dr. Yusuke Tomabechi.

(80) Umekawa, M.; Higashiyama, T.; Koga, Y.; Tanaka, T.; Noguchi, M.; Kobayashi, A.; Shoda, S.; Huang, W.; Wang, L.-X.; Ashida, H. *Biochim. Biophys. Acta - Gen. Subj.* **2010**, *1800* (11), 1203–1209.

(81) Seko, A.; Koketsu, M.; Nishizono, M.; Enoki, Y.; Ibrahim, H. R.; Juneja, L. R.; Kim, M.; Yamamoto, T. *Biochim. Biophys. Acta - Gen. Subj.* **1997**, *1335* (1-2), 23–32.

(82) Davis, B. G.; Lloyd, R. C.; Jones, J. B. *J. Org. Chem.* **1998**, *63* (26), 9614–9615.

(83) Gamblin, D. P.; Garnier, P.; Ward, S. J.; Oldham, N. J.; Fairbanks, A. J.; Davis, B. G. *Org. Biomol. Chem.* **2003**, *1* (21), 3642–3644.

(84) Zhu, X.; Schmidt, R. R. *J. Org. Chem.* **2004**, *69* (4), 1081–1085.

(85) Zhu, X.; Pachamuthu, K.; Schmidt, R. R. *Org. Lett.* **2004**, *6* (7), 1083–1085.

(86) Stowell, C. P.; Lee, Y. C. *Methods Enzymol.* **1982**, *83* (C), 278–288.

(87) MacDougall, J. M.; Zhang, X.-D.; Polgar, W. E.; Khroyan, T. V.; Toll, L.; Cashman, J. R. *J. Med. Chem.* **2004**, *47* (23), 5809–5815.

(88) Gamblin, D. P.; Garnier, P.; van Kasteren, S.; Oldham, N. J.; Fairbanks, A. J.; Davis, B. G. *Angew. Chem. Int. Ed.* **2004**, *116* (7), 846–851.

(89) Luo, S.-Y.; Kulkarni, S. S.; Chou, C.-H.; Liao, W.-M.; Hung, S.-C. *J. Org. Chem.* **2006**, *71* (3), 1226–1229.

(90) Dong, L.; Roosenberg, J. M.; Miller, M. J. *J. Am. Chem. Soc.* **2002**, *124* (50), 15001–15005.

(91) Wang, H.; She, J.; Zhang, L.-H.; Ye, X.-S. *J. Org. Chem.* **2004**, *69* (17), 5774–5777.

(92) Osborn, H. M.; Brome, V. A.; Harwood, L. M.; Suthers, W. G. *Carbohydr. Res.* **2001**, *332* (2), 157–166.

(93) Schelhaas, M.; Waldmann, H. *Angew. Chem. Int. Ed.* **1996**, *35* (18), 2056–2083.

(94) Veeneman, G. H.; Brugghe, H. F.; van den Elst, H.; van Boom, J. H. *Carbohydr. Res.* **1990**, *195* (2), C1–C4.

(95) Zinin, A. I.; Eneyskaya, E. V; Shabalin, K. A; Kulminskaya, A. A; Shishlyannikov, S. M.; Neustroev, K. N. *Carbohydr. Res.* **2002**, *337* (7), 635–

642.

(96) Plusquellec, D.; Roulleau, F.; Bertho, F.; Lefeuvre, M.; Brown, E. *Tetrahedron* **1986**, *42* (9), 2457–2467.

(97) Perrie, J. A.; Harding, J. R.; Holt, D. W.; Johnston, A.; Meath, P.; Stachulski, A. V. *Org. Lett.* **2005**, *7* (13), 2591–2594.

Chapter 6: Experimental Section

General Chemical Experimental

Reactions conducted at 0 °C were cooled by means of an ice bath. Solvent was removed under reduced pressure using a Buchi™ rotary evaporator. Reagents were used as supplied without further purification unless otherwise stated.

Thin Layer Chromatography (t.l.c.) was carried out on Merck Silica Gel $60F_{254}$ aluminium-backed plates. Visualisation of the plates was achieved using a UV lamp (ν_{max} = 254 or 365 nm), and/or ammonium molybdate (5% in 2 M H_2SO_4), and/or aniline-diphenylamine-85 % phosphoric acid (4 mL:4 g:20 mL) in acetone (96 mL), and/or a solution of CuBr (spatula tip amount), ethanol (40 mL), and propargyl alcohol (40 mL). Flash column chromatography was carried out using Sorbisil C60 40/60 silica. Melting points were recorded on an Electrothermal® melting point apparatus. Proton and carbon nuclear magnetic resonance (^1H, ^{13}C) spectra were recorded on Bruker AV400 (400 MHz), and Bruker AV500 (500 MHz) spectrometers. All chemical shifts are quoted on the δ-scale in ppm using residual solvent as an internal standard. ^1H and ^{13}C spectra were assigned using COSY, DEPT, HSQC, HMBC and TOCSY. High resolution mass spectra were recorded a Bruker FT-ICR mass spectrometer using electrospray ionisation (ESI) or chemical ionisation (CI) techniques as stated. *M/z* values are reported in Daltons. Optical rotations were measured on a Perkin-Elmer 241 polarimeter with a water-jacketed 1 cm^3 cell with a path length of 1 dm, and are quoted in units of $°.cm^2.g^{-1}$.

Experimental for Chapter 2

Chemical reagents and equipment suppliers

All chemicals and reagents other than those listed below were supplied by Sigma-Aldrich Chemical Co.

Multiple sequence alignments

Multiple sequence alignments were generated using ClustalW.[1]

Water

All water used in these experiments was from a Millipore Milli-W system. Water used for PCRs was autoclaved prior to use.

pH measurements

The pH of solutions was measured using an UltraBASIC UB-10 Benchtop pH meter (Denver Instrument). Solutions were made more acidic by addition of 1 M or 10 M HCl, and basic by addition of either 1 M or 10 M NaOH.

Plasmids

engEF (pET-23d, ampicillin^r, *C*-$(His)_6$ tagged) provided by Dr. Makoto Ito, Kyushu University, Japan.

engBL (340~1694, pET-28b, kanamycin^r, *N*-$(His)_6$ tagged), provided by Prof. Shinya Fushinobu, University of Tokyo, Japan.

Glycerol stocks

All plasmid-containing strains of *E. coli* were stored frozen at -80 °C as glycerol stocks. Glycerol stocks were created by mixing 0.4 mL of 50 % (v/v) glycerol with 0.8 mL of an overnight culture in a 1.6 mL micro-centrifuge tube.

Antibiotics

The genes for engBL had previously been cloned into the pET-28b vector which carries resistance to kanamycin. Kanamycin was added to all growth media to a final concentration of 0.05 mg/mL. Stock solutions of kanamycin at 50 mg/mL were created, filter sterilized and stored at -80 °C, and freeze-thawed a maximum of two times.

The genes for engEF had previously been cloned into the pET-23d vector which carries resistance to ampicillin. Ampicillin was added to all growth media to a final concentration of 0.1 mg/mL. Stock solutions of ampicillin at 100 mg/mL were created, filter sterilized and stored at -80 °C, and freeze-thawed a maximum of two times.

Site-directed mutagenesis

Mutagenesis was performed using a Quikchange II Site Directed Mutagenesis Kit (Agilent) with Phusion® High-Fidelity DNA Polymerase (Thermo Fisher Scientific). PCRs were performed in a DNA Engine® Peltier Thermal Cycler (Bio-rad). Typically each PCR was 50 μL and used 50-100 ng of plasmid template and 125 ng of each primer. Primers for mutagenesis were synthesised by Invitrogen and resuspended in sterilized Milli-Q water to a concentration of 100 μM. Template DNA (which is methylated) was digested by treatment with the restriction enzyme *DpnI* according to the protocols of kit, Invitrogen.

The forward and reverse primers designed for mutagenesis of engEF were:

D725A Forward 5’ gat tat att tat gtt **gct** gtt tgg ggc aat caa gcc 3’

Reverse 5’ ggc ttg att gcc cca aac **agc** aac ata aat ata atc 3’

D725G Forward 5’ gat tat att tat gtt **ggg** gtt tgg ggc aat caa gcc 3’

Reverse 5’ ggc ttg att gcc cca aac **ccc** aac ata aat ata atc 3’

D725S Forward 5’ gat tat att tat gtt **agc** gtt tgg ggc aat caa gcc 3’

Reverse 5’ ggc ttg att gcc cca aac **gct** aac ata aat ata atc 3’

The forward and reverse primers designed for mutagenesis of engBL were:

D789A Forward 5’ gac ttc atc tac ctc **gct** gtg tgg ggc aac ctg 3’

Reverse 5’ cag gtt gcc cca cac **agc** gag gta gat gaa gtc 3’

D789G Forward 5’ gac ttc atc tac ctc **ggg** gtg tgg ggc aac ctg 3’

Reverse 5’ cag gtt gcc cca cac **ccc** gag gta gat gaa gtc 3’

D789S Forward 5’ gac ttc atc tac ctc **agc** gtg tgg ggc aac ctg 3’

Reverse 5’ cag gtt gcc cca cac **gct** gag gta gat gaa gtc 3’

Plasmid extraction and purification

Plasmids were extracted using a High Pure Plasmid Isolation Kit (Roche). Small cultures (5 mL) of plasmid-containing cells were grown overnight with the appropriate antibiotic. The cultures were harvested the following morning and plasmids extracted and purified as per the kit instructions.

The concentration of purified plasmid was measure by absorption at 260 nm using an appropriately blanked Nanodrop ND-1000 spectrophotometer.

DNA sequencing

All DNA sequencing was performed at Canterbury Sequencing and Genotyping using an Applied Biosystems 3130xl Genetic Analyzer with BigDye Terminator v3.1 (Applied Biosystems) sequencing chemistry. Typically 200 ng to 250 ng of purified double-stranded plasmid and 3.2 μM of primer were supplied for each sample. The primers used for sequencing engEF were:

5’ gacacccacatgtccggcttctgt

5’ ggaagagattttgtcgccttttc

5’ atctgtactattgacagaaatcaa

5’ cacttctggcgtttctccagttgt

5’ atcaattaattgattcaacgcttg

5’ atactctaaagcattcggaaattc

5’ ttgagtaacggtcacttttcagc

5’ atctttaaatggtgttcctgttcc

5’ aaaatcaattcgtatattttgcat

Transformation

Chemically competent cells (10 μL aliquots) were thawed on ice before addition of 2 μL of either PCR product or purified plasmid. The mixture was left on ice for 30 mins then placed in a 42 °C water bath for 30 s to 45 s before being placed on ice again for 2 mins. Cells were then outgrown for 1 h at 37 °C while shaking after addition of 150 μL super optimal broth (SOC) medium.

A portion (50 μL to 200 μL) of the transformed cells was spread directly onto an LB-agar plate containing the appropriate antibiotic and left to grow on the inverted medium overnight at 37 °C.

SOC medium: 2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgSO_4$, 20 mM glucose. Filter sterilized in aliquots.

Cell lines

The cell lines used for plasmid propagation were either *E. coli* TOP10 (Invitrogen) or DH5α. For protein expression, either *E. coli* BL21 (DE3) or BL21 (DE3) Star (Invitrogen) cell lines were used.

Culture media

LB-agar was prepared by dissolving either LB-agar (Miller's) (37 g/L) or LB (Lennox) (20 g/L) and agar (15 g/L) base in Milli-Q water and sterilized by autoclaving. The LB-agar solution was then heated in a microwave oven until boiling, and left to cool in a 60 °C. The LB-agar solution was then further cool to ≤50 °C before antibiotic was added immediately prior to pouring into round petri dishes.

LB media for plasmid extraction, pre-cultures and protein expression cultures were prepared by dissolving 20 g/L LB (Lennox L) base in Milli-Q water and sterilizing by autoclaving. Antibiotic was added immediately prior to use.

Protein expression cultures

Cultures for protein expression were grown in 2 L to 3 L beveled conical flasks in a shaking incubator at 200 rpm. Expression cultures were inoculated with an overnight culture and grown at 37 °C until induction of protein expression.

Transcription of the engEF genes were under the control of a *lac* operon therefore protein expression was induced by addition of IPTG to a final concentration of 0.5 mM, when cultures were at a mid-logarithmic phase (OD_{600} of 0.4 to 0.6 AU). After induction, cultures were kept shaking at 23 °C and harvested the following morning.

Cell harvesting

Typically, large cultures were harvested in 1L bottles at 14 000 *g* for 15 min at 4 °C. The supernatant liquid was removed and the cell pellet either immediately lysed, or stored at -80 °C. Small volume cultures (≤ 5 mL) were progressively harvested in 1.6 mL micro-centrifuge tubes at 16 000 *g* for 1 min.

Cell lysis

Cells were lysed by sonication. Sonication was performed with an Ultrasonic Processor UP200S (Hielscher). Cell pellets were resuspended in chilled suspension buffer (approximately 25 mL) on ice, and sonicated in a beaker surrounded by packed ice. Sonicated at 60 – 70 percent power for 2 minutes intervals with a break of 1 minute

between each interval (5-7 times). The solution was clarified by centrifugation (4 °C, 12,000 rpm, 35 mins). The supernatant was poured off and filtered using a 0.2 μm filter.

Protein purification

All proteins were purified by a two-step procedure composed of, in step order, immobilized metal ion affinity (IMAC), cation exchange chromatography (CEC), and size-exclusion (SEC) chromatography. The chromatography was performed on an ÄKTA FPLC system.

Before use, all buffers were filtered (0.2 μm) under vacuum and degassed. Samples were all likewise filtered before being loaded onto the appropriate columns. The eluates were fractionated in 2 mL fractions collected in 96-well plates. Elution from the columns was monitored principally at 280 nm. Fractions containing protein were then pooled and prepared for the next chromatographic step.

Buffers used at various stages of protein purification

Process	Buffer	Composition
Immobilised metal affinity chromatography	Buffer A	10 mM Tris/HCl, pH 7.5, 150 mM NaCl, 20 mM imidazole
	Buffer B	10 mM Tris/HCl, pH 7.5, 150 mM NaCl, 200 mM imidazole
Ion exchange chromatography	Buffer A	10 mM Tris/HCl, pH 7.5, 20 mM NaCl
	Buffer B	10 mM Tris/HCl, pH 7.5, 1 M NaCl
Size exclusion chromatography	Buffer A	10 mM Tris, 150 mM NaCl, pH 7

Immobilized metal ion affinity chromatography (IMAC)

IMAC was performed using a HisTrap™ FF crude column (5 mL) at 4 °C. The elution from the column was performed using gradients of increasing imidazole (up to 500 mM) at a flow rate of 1 mL/min. The clarified lysate was filtered through a 0.22 µm syringe filter and applied to a HisTrap™ column pre-equilibrated in Buffer A [10 column volumes (CV)] using a Gilson peristaltic pump. The column was washed with 10 CV of buffer A to remove any unbound protein. Pure protein was eluted with a gradient set to 100% buffer B. All flow rates were maintained at 1 mL/min.

Cation exchange chromatography (CEC)

CEC was performed using a Hi-Trap QHB (1 mL) at 4 °C in Buffer A. The elution from the column was performed using gradients of Buffer B (up to 1 M) at a flow rate of 1 mL/min.

Size exclusion chromatography (SEC)

Fractions containing protein from IMAC were pooled and loaded onto a Superdex™ 200 10/300 GL column (GE Healthcare). SEC was performed at 4 °C in Buffer A. The elution flow rate was 0.6 mL/min. The eluted protein fractions were collected as 0.5 mL fractions and relevant fractions were pooled.

Polyacrylamide gel electrophoresis

SDS-PAGE was performed using a Bolt 4-12% Bis-Tris Plus Gel 1.0 mm x 15-well pre-cast gels (Novex®) in NuPAGE® MOPS SDS Running Buffer (Invitrogen). Electrophoresis was performed using a Xcell SureLock™ Electrophoresis Cell

(Invitrogen), at 200 V for 35 mins. Samples were mixed with NuPAGE® LDS Sample Buffer (4x), and boiled, before being loaded onto the gel.

Visualization

All SDS-PAGE gels were stained by SimplyBlue™ Safe Stain (Invitrogen). Gels were first washed and also destained in water, and all steps were performed with heating until almost boiling in a microwave oven. Molecular weight standards were run as the sample in one lane of each gel.

Enzyme storage

Purified protein was divided into aliquots in thin-walled PCR tube in 100 µL aliquots, flash-frozen in liquid nitrogen, and stored at -80 °C. All protein samples were rapidly thawed immediately before use and kept on ice.

Protein concentration determination

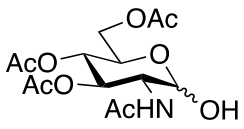
The concentration was measured by absorption at 280 nm by an appropriately blanked Nanodrop ND-1000 spectrophotometer, using 2 µL samples of purified protein solution. The concentration was calculated from the absorption using molar extinction coefficient values of 252910 M^{-1} cm^{-1} and 221510 M^{-1} cm^{-1} for engBL and engEF, respectively. The coefficient values, calculated from protein sequences using the ProtParam tool on the ExPASy Proteomics Server, were the same for both WT and mutant proteins.

CD spectroscopy

The determination of CD spectra were performed on a Jasco J-815 spectropolarimeter in a 2 mm curvette with a 2 second response time. For each experiment, five spectra were summed and averaged to maximize signal-to-noise. Spectra were analysed by Jasco Spectra Manager software (version 1.52).

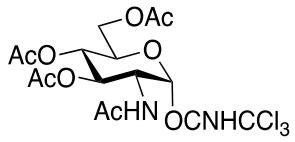
Experimental for Chapter 3

2-Acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-D-glucopyranoside 3.4



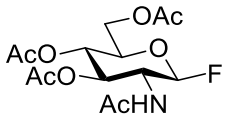
2-Acetamido-1,3,4,6-tetra-*O*-acetyl-2-deoxy-D-glucopyranoside **3.3** (3.1 g, 7.96 mmol) was dissolved in dry DMF (50 mL). Hydrazine acetate (1.1 g, 11.9 mmol) was added and the reaction stirred at rt under an N_2 atmosphere. After 17 h, t.l.c. (ethyl acetate) indicated complete consumption of starting material (R_f 0.4) and formation of a major product (R_f 0.2). The reaction mixture was concentrated *in vacuo* and the residue dissolved in and extracted with ethyl acetate (5 x 100 mL), washed with brine (100 mL), dried ($MgSO_4$), filtered and concentrated *in vacuo*. The residue was purified by flash column chromatography (ethyl acetate) to give 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-D-glucopyranoside **3.4** (2.0 g, 72 %) as a pale oil; δ_H (400 MHz, $CDCl_3$)[2] (only α anomer quoted) 1.96, 2.02, 2.03, 2.09 (12H, 4 x s, 4 x CH_3CO_2), 4.11-4.21 (3H, m, H-5, H-6, H-6'), 4.29 (1H, m, H-2), 5.10-5.32 (3H, m, H-1, H-3, H-4), 5.90 (1H, d, $J_{2,NH}$ 9 Hz, N*H*); HMRS (ESI) Calcd. For $C_{14}H_{21}NO_9$ (MNa^+) 370.1114. Found 370.1109.

2-Acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-α-D-glucopyranosyl trichloroacetimidate 3.5



2-Acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-D-glucopyranoside **3.4** (290 mg, 0.835 mmol) and 1,1,1-trichloroacetonitrile (0.9 mL, 8.35 mmol) were stirred in dry DCM (10 mL) and cooled to 0 °C. DBU (25 μL, 0.288 mmol) was added. After 3 h, t.l.c. (ethyl acetate) indicated complete consumption of starting material (R_f 0.2) and the formation of a major product (R_f 0.7). The reaction mixture was concentrated, and the residue purified by flash column chromatography (ethyl acetate) to give 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-α-D-glucopyranosyl trichloroacetimidate **3.5** (380 mg, 93 %) as a yellow oil; δ_H (400 MHz, $CDCl_3$)[2] 1.85, 1.97, 1.99 (12H, 3 x s, 4 x CH_3CO_2), 4.01-4.05 (2H, m, H-5, H-6), 4.15 (1H, m, H-6'), 4.45 (1H, m, H-2), 5.12-5.26 (2H, m, H-3, H-4), 5.77 (1H, d, $J_{2,NH}$ 9.0 Hz, N*H*CO_2) 6.28 (1H, d, $J_{1,2}$ 2.3 Hz, H-1), 8.77 (1H, s, N*H*); HRMS (ESI) Calcd. For $C_{16}H_{21}Cl_3N_2O_9$ (MNa^+) 513.0210. Found 513.0210.

2-Acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-β-D-glucopyranosyl fluoride 3.6



2-Acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-D-glucopyranosyl trichloroacetimidate **3.5** (380 mg, 0.775 mmol) was stirred in dry DCM (20 mL) in a Teflon tube at 0 °C. HF

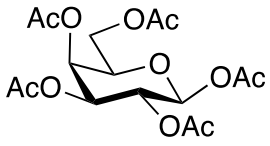
(70% solution in pyridine, 100 μL) was added. After 5 mins, t.l.c. (ethyl acetate) indicated complete consumption of starting material (R_f 0.7) and the formation of a major product (R_f 0.6). The reaction mixture was diluted with DCM (100 mL), washed with potassium hydrogen carbonate (100 mL of a saturated solution), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. The residue was purified by flash column chromatography (ethyl acetate/petrol, 1:1) and recrystallized (ethyl acetate/petrol) to give 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-β-D-glucopyranosyl fluoride **3.6** (90 mg, 37 %) as a white crystalline solid, m.p. 130 °C (ethyl acetate/petrol); $[\alpha]_D^{20}$ +16 (*c*, 1.0 in $CHCl_3$); ν_{max} (KBr disc) 3258 (br, NH) cm^{-1}, 1744 (s, C=O) cm^{-1}; δ_H (400 MHz, $CDCl_3$) 1.99, 2.05, 2.06, 2.10 (12H, 4 x s, 4 x CH_3CO_2), 3.89 (1H, m, H-5), 4.12 (1H, m, H-2), 4.25 (2H, m, H-6, H-6'), 5.14 (1H, m, H-4), 5.23 (1H, m, H-3), 5.42 (1H, dd, $J_{1,2}$ 6.3 Hz, $J_{F,1}$ 51.7 Hz, H-1) 5.79 (1H, d, $J_{2,NH}$ 7.5 Hz, N*H*); δ_C (100.5 MHz, $CDCl_3$) 20.6, 20.6, 20.7, 23.1 (2 x s, 4 x C(O)$\underline{C}H_3$), 53.8 (d, C-5), 62.0 (d, C-6), 67.7 (s, C-4), 70.7 (d, C-3), 72.1 (d, C-5), 106.0 (d, C-1), 169.2, 170.6, 170.6, 170.6 (4 x s, 4 x C=O); δ_F (376.23 MHz, $CDCl_3$, CF_3COOH standard) -147.6 (dd, $J_{F,1}$ 51.8 Hz, $J_{F,2}$ 25.9 Hz, C-F); HRMS (ESI) Calcd. For $C_{14}H_{20}FNO_8$ (MNa^+) 372.1071. Found 372.1064.

2-Acetamido-2-deoxy-β-D-galactopyranosyl azide 3.13[3]

N-Acetyl-D-galactosamine **3.10** (5 g, 22.6 mmol) and triethylamine (15.6 mL, 113.0 mmol) were stirred in H_2O/MeCN (4:1, 100 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **3.11** (19.3 g, 67.8 mmol) was added.

After 3 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.2) and the formation of two major products (R_f 0.5, 0.6). The reaction mixture was acidified to pH 2 by dropwise addition of aqueous 1.2 M HCl and then neutralized by addition of sat. $NaHCO_3$ solution (100 mL). The reaction mixture was concentrated, and the residue dissolved in ethanol (20 mL), filtered through Celite®, and concentrated *in vacuo*. The residue was then redissolved in water (50 mL), washed with DCM (2 x 50 mL), filtered through a column of Amberlite® IR120 (H^+, previously treated with 1 M NaOH solution), and concentrated *in vacuo*. Purification by flash column chromatography ($CHCl_3$:MeOH, 7:1 → $CHCl_3$:MeOH, 2:1) gave 2-acetamido-2-deoxy-β-D-galactopyranosyl azide **3.13** (4.3 g, 77 %) as a white solid; m.p. 195-200 °C; $[\alpha]_D^{20}$ -16 (*c*, 1.0 in MeOH); ν_{max} (neat) 1643 (s, amide), 2094 (s, azide), 3331.3 (bs, OH) cm^{-1}; δ_H (400 MHz, D_2O) 1.94 (3H, s, CH_3), 3.63-3.78 (4H, m, H-3, H-5, H-6, H-6'), 3.80 (1H, at, *J* 10.2 Hz, H-2), 3.86 (1H, ad, *J* 3.1 Hz, H-4), 4.55 (1H, d, $J_{1,2}$ 9.4 Hz, H-1); δ_C (100 MHz, D_2O) 22.1 (d, NHCO<u>C</u>H_3), 51.7 (d, C-2), 60.9 (t, C-6), 67.6 (d, C-4), 70.7 (d, C-3), 77.2 (d, C-5), 89.0 (d, C-1), 175.0 (s, NH<u>C</u>OCH_3); HRMS (ESI) Calcd. For $C_8H_{14}N_4O_5$ (MNa^+) 269.0862. Found 269.0854.

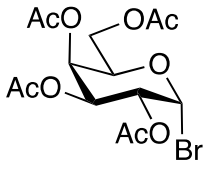
1,2,3,4,6-Penta-*O*-acetyl-β-D-galactopyranoside 3.15



Sodium acetate (82.0 g, 1000 mmol) was added to acetic anhydride (630 mL) and refluxed at 120 °C for 1 h. D-Galactose **3.14** (60 g, 330 mmol) was then added portion-wise while the reaction was vigorously stirred over a period of 4 h, at which time t.l.c.

(petrol:ethyl acetate, 1:1) indicated complete consumption of starting material (R_f 0.1) and formation of a single product (R_f 0.7). The reaction mixture was allowed to cool to rt, and then poured into ice-cold water (10 L). The resulting precipitate was filtered, and washed with water. Filtration and purification of the crude residue by recrystallization (ethanol) to afford 1,2,3,4,6-penta-*O*-acetyl-β-D-galactopyranoside **3.2** (48 g, 37 %) as a white crystalline solid, m.p. 125 °C [lit. m.p. 137-139 °C];[4] $[\alpha]_D^{20}$ +29.1 (*c*, 1.0 in $CHCl_3$) [lit. $[\alpha]_D^{23.5}$ +27.1 (*c*, 1.03 in $CHCl_3$)];[5] δ_H (500 MHz, $CDCl_3$)[6] 1.99, 2.04, 2.11, 2.16 (15H, 4 x s, 5 x CH_3CO_2), 4.07 (1H, m, H-5), 4.13-4.16 (2H, m, H-6, H-6'), 5.06 (1H, dd, $J_{3,4}$ 3.5 Hz, H-3), 5.32 (1H, dd, $J_{2,3}$ 10.5 Hz, H-2), 5.42 (1H, m, H-4), 5.69 (1H, d, $J_{1,2}$ 8.3 Hz, H-1); HRMS (ESI) Calcd. For $C_{16}H_{22}O_{11}$ (MNa^+) 413.1054. Found 413.1054.

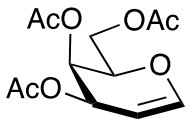
2,3,4,6-Tetra-*O*-acetyl-α-D-galactopyranosyl bromide 3.16[7]



1,2,3,4,6-Penta-*O*-acetyl-β-D-galactopyranoside **3.15** (50 g, 128 mmol) was dissolved in DCM (450 mL) and cooled to 0 °C. Hydrogen bromide solution (33 % in AcOH solution, 140 mL) was added slowly and the reaction was stirred at rt. After 3 h, t.l.c. (petrol:ethyl acetate, 2:1) indicated the complete consumption of starting material (R_f 0.3) and formation of a single product (R_f 0.5). The reaction mixture was transferred to ice-water (500 mL) and the organic layer was extracted with DCM (100 mL), washed with sodium hydrogen carbonate (4 x 1 L of a saturated solution), brine (500 mL), dried

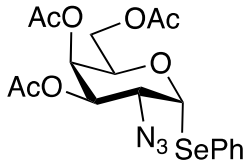
($MgSO_4$), filtered, and concentrated *in vacuo* to yield 2,3,4,6-tetra-*O*-acetyl-α-D-galactopyranosyl bromide **3.16** (52.3 g, 100 %) as a yellow oil; δ_H (500 MHz, $CDCl_3$)[8] 1.90, 1.95, 2.01, 2.05 (12H, 4 x s, CH_3CO_2), 3.99-4.11 (2H, m, $J_{5,6'}$ 6.5 Hz, $J_{6,6'}$ 11.4 Hz, H-6, H-6'), 4.40 (1H, at, H-5), 4.94 (1H, dd, $J_{1,2}$ 3.9 Hz, $J_{2,3}$ 10.6 Hz, H-2), 5.29 (1H, dd, $J_{3,4}$ 3.1 Hz, $J_{4,5}$ 10.6 Hz, H-4), 5.41 (1H, ad, H-3), 6.61 (1H, d, H-1). This material was used without further purification or characterisation.

3,4,6-Tri-*O*-acetyl-D-galactal 3.17[8]



Activated zinc dust (50.7 g, 769 mmol) and *N*-methylimidazole (10.2 mL, 128 mmol) was stirred in ethyl acetate (200 mL) and the mixture was refluxed. Galactosyl bromide **3.16** (52.3 g, 128 mmol) dissolved in ethyl acetate (200 mL) was added drop-wise to the stirring solution. After 2 h, the reaction mixture was filtered through Celite® washed with ethyl acetate (100 mL). The filtrate was washed with hydrochloric acid (100 mL of a 1.2 M solution), sodium hydrogen carbonate (2 x 200 mL of a saturated solution), dried ($MgSO_4$), and concentrated *in vacuo*. The residue was purified by flash column chromatography (petrol:ethyl acetate, 2:1) to yield 3,4,6-tri-*O*-acetyl-D-galactal **3.17** (27.6 g, 79 %) as a pale yellow oil, $[\alpha]_D^{20}$ -9 (*c*, 1.0 in EtOAc) [lit. $[\alpha]_D^{25}$ -9 (*c*, 1.0 in EtOAc)];[9] δ_H (500 MHz, $CDCl_3$)[8] 2.01, 2.07, 2.12 (9H, 3 x s, 3 x CH_3CO_2), 4.20-4.31 (3H, m, H-5, H-6, H-6'), 4.71 (1H, dd, $J_{1,2}$ 6.2 Hz, $J_{2,3}$ 1.5 Hz, H-2), 5.41 (1H, m, H-4), 5.53 (1H, m, H-3), 6.44 (1H, dd, H-1); HRMS (ESI) Calcd. For $C_{12}H_{16}O_7$ (MNa^+) 295.0788. Found 295.0796.

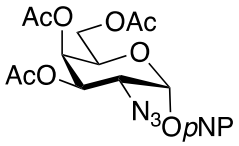
Phenyl 2-azido-2-deoxy-3,4,6-tri-*O*-acetyl-1-seleno-α-D-galactopyranoside 3.18[10]



Tri-*O*-acetyl-D-galactal **3.17** (8.3 g, 30.5 mmol) was dissolved in dry DCM (50 mL) and stirred at rt under an N_2 atmosphere. A solution containing bis(acetoxy)iodobenzene (13.8 g, 42.7 mmol) and diphenyl diselenide (5.7 g, 18.3 mmol) dissolved in dry DCM (50 mL) was added, and the reaction mixture cooled to 0 °C. Sodium azide (4.8 g, 73.2 mmol) was then added slowly and the reaction mixture was allowed to warm to rt. After 17 h, t.l.c. (petrol:ethyl acetate, 3:1) indicated the complete consumption of starting material (R_f 0.2) and formation of two major products (R_f 0.3, 0.35). The reaction mixture was diluted with DCM (300 mL), washed with water (4 x 250 mL), sodium hydrogen carbonate (2 x 100 mL of a saturated solution), brine (2 x 100 mL), dried ($MgSO_4$), and concentrated *in vacuo*. The residue was purified by flash column chromatography (petrol:ethyl acetate, 4:1) and recrystallized (petrol/ethyl acetate) to yield phenyl 2-azido-2-deoxy-3,4,6-tri-*O*-acetyl-1-seleno-α-D-galactopyranoside **3.18** (5.7 g, 40 %) as a white crystalline solid, m.p. 100 °C [lit. m.p. 104-105 °C];[11] $[\alpha]_D^{20}$ +156 (*c*, 1.0 in EtOAc) [lit. $[\alpha]_D$ + 156 (*c*, 1.03 in EtOAc)];[11] δ_H (500 MHz, $CDCl_3$)[11] 1.99, 2.08, 2.17 (9H, 3 x s, 3 x CH_3CO_2), 4.02-4.10 (2H, m, $J_{5,6}$ 5.7 Hz, $J_{6,6'}$ 11.4 Hz, H-6, H-6'), 4.27 (1H, dd, $J_{1,2}$ 5.5 Hz, $J_{2,3}$ 11 Hz, H-2), 4.69 (1H, m, H-5), 5.13 (1H, dd, $J_{3,4}$ 3.3 Hz, H-3), 5.49 (1H, at, H-4), 6.02 (1H, d, H-1), 7.29-

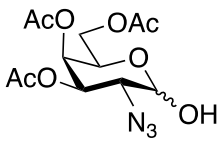
7.63 (5H, m, Ar-H); HRMS (ESI) Calcd. For $C_{18}H_{21}N_3O_7Se$ (MNa^+) 494.0437. Found 494.0440.

_p_-Nitrophenyl 3,4,6-tri-_O_-acetyl-2-azido-2-deoxy-α-D-galactopyranoside 3.19



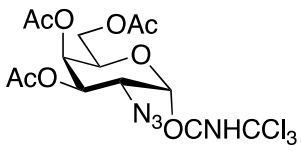
3,4,6-Tri-*O*-acetyl-2-azido-2-deoxy-α-D-galactopyranosyl trichloroacetimidate **3.21** (1.8 g, 3.79 mmol), *p*-nitrophenol (634 mg, 4.56 mmol) and pre-dried 4Å molecular sieves (5 g) in dry DCM (40 mL) were stirred at -20 °C under an N_2 atmosphere. TMSOTf (0.69 mL, 3.79 mmol) was added dropwise. After 3.5 h, t.l.c. (petrol:ethyl acetate, 2:1) indicated complete consumption of starting material (R_f 0.4) and formation of a major product (R_f 0.3). The reaction mixture was diluted with DCM (100 mL) and the organic phase was washed with sodium hydrogen carbonate (100 mL of a saturated solution), brine (100 mL), dried ($MgSO_4$), and concentrated *in vacuo*. The crude product was purified by flash column chromatography (petrol:ethyl acetate, 4:1) to give *p*-nitrophenyl 3,4,6-tri-*O*-acetyl-2-azido-2-deoxy-α-D-galactopyranoside **3.19** (1.3 g, 76 %) as a white solid, m.p. 110-113 °C; δ_H (400 MHz, $CDCl_3$)[12] 1.95, 2.11, 2.19 (9 H, 3 x s, 3 x CH_3CO_2), 3.91 (1H, dd, $J_{1,2}$ 2.7 Hz, $J_{2,3}$ 10.9 Hz, H-2), 4.00-4.13 (2H, m, H-6, H-6'), 4.26 (1H, m, H-5), 5.03 (1H, at, *J* 10.2 Hz, H-4), 5.55 (1H, m, H-3), 5.74 (1H, m, H-1), 7.14-7.23 (2H, m, Ar-H), 8.23-8.26 (2H, m, Ar-H); HRMS (ESI) Calcd. For $C_{18}H_{20}N_4O_{10}$ (MNa^+) 475.1072. Found 475.0176.

3,4,6-Tri-*O*-acetyl-2-azido-2-deoxy-D-galactopyranoside 3.20[11]



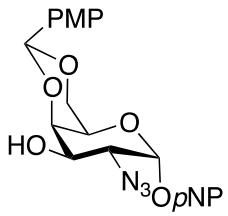
Phenyl 2-azido-2-deoxy-3,4,6-tri-*O*-acetyl-1-seleno-α-D-galactopyranoside **3.18** (5 g, 10.0 mmol) in THF-H_2O (50 mL, 1:1) was treated at rt with *N*-iodosuccinimide (7.2 g, 31.8 mmol). After 5 h, t.l.c. (petrol:ethyl acetate, 2:1) indicated complete consumption of starting material (R_f 0.5) and the formation of a major product (R_f 0.2). The mixture was diluted with ethyl acetate (100 mL) and washed with brine (100 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. The crude product was purified by flash column chromatography (petrol:ethyl acetate, 3:1) to give 3,4,6-tri-*O*-acetyl-2-azido-2-deoxy-D-galactopyranoside **3.20** (3.5 g, 99 %) as a pale yellow oil; δ_H (500 MHz, $CDCl_3$)[13] 2.04-2.16 (9H, 3 x s, 3 x CH_3CO_2), 3.52 (1H, d, $J_{1,OH}$ 2.4 Hz, OH), 3.66 (1H, dd, $J_{1,2}$ 8.1 Hz, $J_{2,3}$ 10.8 Hz, H-2β), 3.74 (1H, dd, $J_{1,2}$ 3.1 Hz, $J_{2,3}$ 11.0 Hz, H-2α), 3.90 (1H, m, H-5β), 4.06-4.14 (4H, m, H-6α, H-6'α, H-6β, H-6'β), 4.45 (1H, at, H-5α), 4.70 (1H, dd, $J_{1,OH}$ 5.1 Hz, $J_{1,2}$ 7.5 Hz, H-1β), 4.81 (1H, dd, $J_{2,3}$ 11.0 Hz, $J_{3,4}$ 3.3 Hz, H-3β), 5.33-5.39 (2H, m, H-4α, H-4β), 5.40 (1H, m, H-3α), 5.46 (1H, d, H-1α); HRMS (ESI) Calcd. For $C_{12}H_{17}N_3O_8$ (MNa^+) 354.0908. Found 354.0911.

3,4,6-Tri-*O*-acetyl-2-azido-2-deoxy-α-D-galactopyranosyl trichloroacetimidate 3.21[14]



3,4,6-Tri-*O*-acetyl-2-azido-2-deoxy-D-galactopyranoside **3.20** (1.34 g, 4.05 mmol) and 1,1,1-trichloroacetonitrile (4.1 mL, 40.5 mmol) were dissolved in dry DCM (40 mL) and the mixture was stirred at 0 °C. DBU (0.12 mL, 0.81 mmol) was added. After 3 h, t.l.c. (petrol:ethyl acetate, 2:1) indicated complete consumption of starting material (R_f 0.2) and the formation of a major product (R_f 0.4). The reaction mixture was concentrated, and purified by flash column chromatography (petrol:ethyl acetate, 3:1) to give 3,4,6-tri-*O*-acetyl-2-azido-2-deoxy-α-D-galactopyranosyl trichloroacetimidate **3.21** (1.88 g, 98 %) as a yellow oil; $[\alpha]_D^{20}$ +102.6 (*c*, 1.0 in $CHCl_3$) [lit. $[\alpha]_D^{25}$ +104 (*c*, 1.0 in $CHCl_3$);[15] δ_H (400 MHz, $CDCl_3$)[15] 2.01, 2.08, 2.17 (9H, 3 x s, 3 x CH_3CO_2), 4.01-4.18 (3H, m, H-2, H-6, H-6'), 4.42 (1H, m, H-5), 5.38 (1H, dd, $J_{2,3}$ 11.0 Hz, $J_{3,4}$ 3.2 Hz, H-3), 5.54 (1H, m, H-4), 6.51 (1H, d, $J_{1,2}$ 3.5 Hz, H-1), 8.79 (1H, s, NH); HRMS (ESI) Calcd. For $C_{14}H_{17}O_8N_4Cl_3$ (MNa^+) 497.0004. Found 497.0021.

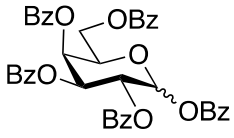
_p_-Nitrophenyl 2-azido-2-deoxy-4,6-_O_-_p_-methoxybenzylidene-α-D-galactopyranoside 3.22



_p_-Nitrophenyl 3,4,6-tri-_O_-acetyl-2-azido-2-deoxy-α-D-galactopyranoside **3.19** (1.2 g, 2.65 mmol) was dissolved in dry methanol (40 mL). Sodium metal (920 mg, 39.2 mmol) was then added portionwise. After 2 h, t.l.c. (petrol:ethyl acetate, 2:1) indicated complete consumption of starting material (R_f 0.3) and the formation of a single product (R_f 0). The reaction mixture was neutralized using Amberlite® IR120 (H^+), filtered and concentrated *in vacuo*. The residue was dissolved in acetonitrile (20 mL) and DMF (8 mL). Anisaldehyde dimethyl acetal (1.4 mL, 7.95 mmol) and *p*-toluene sulfonic acid (250 mg, 1.33 mmol) were added, and the reaction was stirred at rt under an N_2 atmosphere. After 24 h, t.l.c. (toluene:ethyl acetate, 2:1) indicated complete consumption of starting material (R_f 0) and the formation of a major product (R_f 0.5). The reaction mixture was diluted with DCM (100 mL) and the organic phase was separated and washed with sodium hydrogen carbonate (200 mL), brine (200 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. The crude product was purified by flash column chromatography (toluene:ethyl acetate, 5:1) to give *p*-nitrophenyl 2-azido-2-deoxy-4,6-*O*-*p*-methoxybenzylidene-α-D-galactopyranoside **3.22** (440 mg, 37 %) as a white solid, m.p. 150 °C; $[\alpha]_D^{20}$ +66.7 (*c*, 1.7 in $CHCl_3$) [lit. $[\alpha]_D^{20}$ +68 (*c*. 1.7 in $CHCl_3$)];[16] δ_H (400 MHz, $CHCl_3$)[16] 2.55 (1H, bs, OH), 3.79-3.84 (5H, m, $C_6H_4OCH_3$, H-2, H-5), 4.04-4.08 (1H, m, H-6'), 4.26 (1H, m, H-6), 4.38 (2H, m, H-3, H-4), 5.58 (1H, s, $CHC_6H_4OCH_3$), 5.78 (1H, d, $J_{1,2}$ 2.7 Hz, H-1), 6.91-6.94 (2H, m, $C_6H_4OCH_3$),

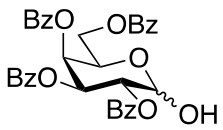
7.20-7.23 (2H, m, $OC_6H_4NO_2$), 7.42-7.45 (2H, m, $OC_6H_4NO_2$), 8.22-8.25 (2H, m, $OC_6H_4NO_2$); HRMS (ESI) Calcd. For $C_{20}H_{20}N_4O_8$ (MNa^+) 467.1173. Found 467.1177.

1,2,3,4,6-Penta-*O*-benzoyl-D-galactopyranoside 3.23[17]



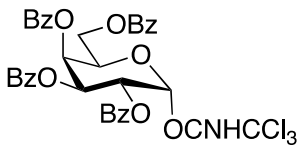
D-Galactose **3.14** (5 g, 27.8 mmol) was stirred in pyridine (150 mL) at 0 °C under an N_2 atmosphere. Benzoyl chloride (21 mL, 172 mmol) was added dropwise and the reaction was allowed to warm to rt. After 18 h, t.l.c. (petrol:ethyl acetate, 2:1) indicated complete consumption of starting material (R_f 0) and the formation of a single product (R_f 0.5). The reaction mixture was diluted with ethyl acetate (500 mL), washed with hydrochloric acid (8 x 250 mL of a 1.2 M solution), sodium hydrogen carbonate (4 x 200 mL of a saturated solution), brine (2 x 200 mL), dried ($MgSO_4$), and concentrated *in vacuo*. Recrystallisation (petrol:ethyl acetate) gave 1,2,3,4,6-penta-*O*-benzoyl-D-galactopyranoside **3.23** (15.3 g, 79 %) as a white solid, m.p. 130 °C [lit. m.p. 129 -131 °C];[17] δ_H (500 MHz, $CDCl_3$)[17] (major α anomer quoted) 4.42 (1H, dd, $J_{5,6}$ 7.1 Hz, $J_{6,6'}$ 11.2 Hz, H-6) 4.63 (1H, dd, $J_{5,6'}$ 6.4 Hz, H-6'), 4.83 (1H, at, H-5), 6.02 (1H, dd, $J_{1,2}$ 3.7 Hz, $J_{2,3}$ 10.8 Hz, H-2), 6.12 (1H, at, *J* 10.5 Hz) H-3), 6.19 (1H, m, H-4), 6.95 (1H, d, H-1), 7.27-8.13 (25H, m, Ar-H); HRMS (ESI) Calcd. For $C_{41}H_{32}O_{11}$ (MNa^+) 723.1837. Found 723.1847.

2,3,4,6-Tetra-*O*-benzoyl-D-galactopyranoside 3.24



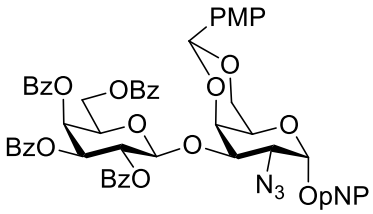
1,2,3,4,6-Penta-*O*-benzoyl-D-galactopyranoside **3.23** (5 g, 7.14 mmol) was dissolved in dry DMF (50 mL). Hydrazine acetate (990 mg, 10.7 mmol) was added and the reaction was stirred at rt. After 5.5 h, t.l.c. (petrol:ethyl acetate, 2:1) indicated complete consumption of starting material (R_f 0.5) and the formation of a major product (R_f 0.3). The reaction mixture was diluted with ethyl acetate (200 mL) and the organic phase was washed with water (4 x 100 mL), sodium hydrogen carbonate (2 x 100 mL of a saturated solution), dried ($MgSO_4$), and concentrated *in vacuo*. The residue was purified by flash column chromatography (petrol:ethyl acetate, 3:1) to give 2,3,4,6-tetra-*O*-benzoyl-D-galactopyranoside **3.24** (4.05 g, 95 %) as a white solid, m.p. 75 °C; δ_H (400 MHz, $CDCl_3$)[17] (only α anomer quoted) 4.37-4.42 (1H, m, H-6'), 4.59-4.64 (1H, m, H-6), 4.87 (1H, m, H-5), 5.71 (1H, at, *J* 3.5 Hz, H-2), 5.85 (1H, d, $J_{1,2}$ 3.5 Hz, H-1), 6.07 (2H, m, H-3, H-4), 7.22-8.10 (20H, m, Ar-H); HRMS (ESI) Calcd. For $C_{34}H_{28}O_{10}$ (MNa^+) 619.1575. Found 619.1599.

2,3,4,6-Tetra-*O*-benzoyl-α-D-galactopyranosyl trichloroacetimidate 3.25



2,3,4,6-Tetra-*O*-benzoyl-D-galactopyranoside **3.24** (1 g, 1.68 mmol) and 1,1,1-trichloroacetonitrile (1.7 mL, 16.8 mmol) were stirred in dry DCM (20 mL) and cooled to 0 °C. DBU (50 μL, 0.34 mmol) was added. After 5 h, t.l.c. (petrol:ethyl acetate, 2:1) indicated complete consumption of starting material (R_f 0.3) and the formation of a major product (R_f 0.6). The reaction mixture was concentrated, and the residue purified by flash column chromatography (petrol:ethyl acetate, 5:1) to give 2,3,4,6-tetra-*O*-benzoyl-α-D-galactopyranosyl trichloroacetimidate **3.25** (1.0 g, 81 %) as a white solid; $[\alpha]_D^{20}$ +99 (*c*, 6.0 in $CHCl_3$) [lit. $[\alpha]_D$ +105 (*c*, 6.05 in $CHCl_3$)];[17] δ_H (400 MHz, $CDCl_3$)[17] 4.44 (1H, dd, $J_{5,6}$ 5.9 Hz, $J_{6,6'}$ 11.3 Hz, H-6'), 4.61 (1H, dd, $J_{5,6}$ 7.0 Hz, H-6), 4.86 (1H, at, *J* 6.3 Hz, H-5), 5.96 (1H, dd, $J_{1,2}$ 3.9 Hz, $J_{2,3}$ 10.9 Hz, H-2), 6.07 (1H, dd, $J_{3,4}$ 3.1 Hz, H-3), 6.15 (1H, m, H-5), 6.92 (1H, d, H-1), 7.24-8.10 (20H, m, Ar-H), 8.63 (1H, s, N*H*); HRMS (ESI) Calcd. For $C_{36}H_{28}Cl_3NO_{10}$ (MNa^+) 762.0671. Found 762.0673.

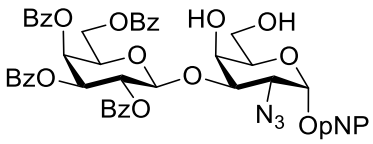
_p_-Nitrophenyl 2,3,4,6-tetra-_O_-benzoyl-β-D-galactopyranosyl-(1→3)-2-azido-2-deoxy-4,6-_O_-_p_-methoxybenzylidene-α-D-galactopyranoside 3.26[16]



A mixture of *p*-nitrophenyl 2-azido-2-deoxy-4,6-*O*-*p*-methoxybenzylidene-α-D-galactopyranoside **3.22** (300 mg, 0.68 mmol), 2,3,4,6-tetra-*O*-benzoyl-α-D-galactopyranosyl trichloroacetimidate **3.25** (600 mg, 0.81 mmol) and 4Å MS in dry DCM (20 mL) was stirred for 30 min under an N_2 atmosphere and then cooled to -20 °C. TMSOTf (6.1 μL, 0.034 mmol) was then added and the reaction mixture was stirred at -20 °C. After 4 h, t.l.c. (toluene:ethyl acetate, 7:1) indicated complete consumption of acceptor **3.22** (R_f 0.2) and the formation of a major product (R_f 0.4). The mixture was diluted with DCM (100 mL) and filtered through a pad of Celite® The filtrate was washed with hydrochloric acid (50 mL of a 1.2 M solution), sodium hydrogen carbonate (50 mL of a saturated solution), water (100 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. Purification by flash column chromatography (toluene:ethyl acetate, 20:1) afforded *p*-nitrophenyl 2,3,4,6-tetra-*O*-benzoyl-β-D-galactopyranosyl-(1→3)-2-azido-2-deoxy-4,6-*O*-*p*-methoxybenzylidene-α-D-galactopyranoside **3.26** (310 mg, 45 %) as a white solid, m.p. 90 °C; $[\alpha]_D^{20}$ +150 (*c*. 1.7 in $CHCl_3$) [lit. $[\alpha]_D^{20}$ +156 (*c*, 1.63 in $CHCl_3$)][16] δ_H (400 MHz, $CDCl_3$)[16] 3.50 (1H, m, H-5), 3.69-3.79 (1H, m, H-6'b), 3.82 (3H, s, $C_6H_4OC\mathit{H_3}$), 4.02-4.11 (2H, m, H-6a, H-2a), 4.34 (1H, dd, $J_{2a,3a}$ 10.6 Hz, $J_{3a,4a}$ 3.2 Hz, H-3a), 4.40-4.45 (2H, m, H-5b, H-6'b), 4.80 (1H, dd, $J_{6b,6b'}$ 12.5 Hz, $J_{5b,6b}$ 8.6 Hz, H-6b), 5.25 (1H, d, $J_{1b,2b}$ 7.8 Hz, H-1b), 5.48 (1H, s, $C\mathit{H}C_6H_4OCH_3$), 5.63 (1H, dd, $J_{2b,3b}$ 10.2 Hz, $J_{3b,4b}$ 3.1 Hz, H-3b), 5.72 (1H, d, $J_{1a,2a}$ 3.1 Hz, H-1a), 5.95

(1H, dd, $J_{1b,2b}$ 7.8 Hz, $J_{2,3}$ 10.2 Hz, H-2b), 6.02 (1H, m, H-4b), 6.87-6.90 (2H, m, $C_6H_4OCH_3$), 7.13-7.19 (2H, m, $OC_6H_4NO_2$), 7.33-7.39 (4H, m, C_6H_5), 7.40-7.49 (6H, m, C_6H_5, $C_6H_4OCH_3$), 7.50-7.53 (1H, m, C_6H_5), 7.60-7.65 (1H, m, C_6H_5), 7.79-7.82 (2H, m, C_6H_5), 7.97-7.99 (2H, m, C_6H_5), 8.00-8.03 (2H, m, C_6H_5), 8.08-8.11 (2H, m, C_6C_5), 8.19-8.21 (2H, m, $OC_6H_4NO_2$); HRMS (ESI) Calcd. For $C_{54}H_{47}N_4O_{17}$ (MH^+) 1045.2750. Found 1045.2728.

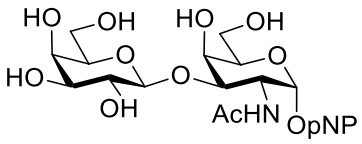
_p_-Nitrophenyl 2,3,4,6-tetra-_O_-benzoyl-β-D-galactopyranosyl-(1→3)-2-azido-2-deoxy-α-D-galactopyranoside 3.27[16]



Acetic acid (80 % aq. solution, 20 mL) was added to a solution of *p*-nitrophenyl 2,3,4,6-tetra-*O*-benzoyl-β-D-galactopyranosyl-(1→3)-2-azido-2-deoxy-4,6-*O*-*p*-methoxybenzylidene-α-D-galactopyranoside **3.26** (200 mg, 0.196 mmol) in DCM (4 mL) and stirred at rt overnight. After 19 h, t.l.c. (toluene:ethyl acetate, 2:1) indicated complete consumption of starting material (R_f 0.8) and the formation of a major product (R_f 0.3). The mixture was diluted with DCM (50 mL) and carefully washed with sodium hydrogen carbonate (2 x 100 mL of saturated solution), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. The crude product was purified by flash column chromatography (toluene:ethyl acetate, 2:1) to afford *p*-nitrophenyl 2,3,4,6-tetra-*O*-benzoyl-β-D-galactopyranosyl-(1→3)-2-azido-2-deoxy-α-D-galactopyranoside **3.27** (150 mg, 85 %) as a white solid, m.p. 100 °C; $[\alpha]_D^{20}$ +175 (*c*, 0.1 in $CHCl_3$) [lit. $[\alpha]_D^{20}$ +167 (*c*, 0.15 in $CHCl_3$)];[16] δ_H (400 MHz, $CHCl_3$)[16] 2.16 (1H, bs, OH), 3.00 (1H, bs, OH), 3.57-3.62 (1H, m, H-6'a), 3.69-3.74 (1H, m, H-6a), 3.75-3.77 (1H, m, H-5a),

3.87 (1H, dd, $J_{2a,3a}$ 10.4 Hz, $J_{3a,4a}$ 2.5 Hz, H-2a), 4.27 (1H, ad, H-3a), 4.36 (1H, m, H-4a), 4.48-4.51 (1H, m, H-5b) 4.54 (1H, dd, $J_{6b,6b'}$ 11.7 Hz, $J_{5b,6b'}$ 3.9 Hz, H-6'b), 4.68 (1H, dd, $J_{6b,6b'}$ 11.3 Hz, $J_{5b,6b'}$ 7.4 Hz, H-6b), 5.15 (1H, d, $J_{1b,2b}$ 7.8 Hz, H-1b), 5.64-5.73 (2H, m, H-3b, H-1a), 5.94 (1H, at, H-2b), 6.05 (1H, m, H-4b), 7.08-7.11 (2H, m, $OC_6H_4NO_2$), 7.21-7.25 (2H, m, C_6H_5), 7.32-7.44 (5H, m, C_6H_5), 7.46-7.55 (4H, m, C_6H_5), 7.59-7.64 (1H, m, C_6H_5), 7.77-7.80 (2H, m, C_6H_5), 7.99-8.01 (4H, m, C_6H_5), 8.09-8.15 (4H, m, $OC_6H_4NO_2$, C_6H_5); HRMS (ESI) Calcd. For $C_{46}H_{40}N_4O_{16}$ (MNa^+) 927.2332. Found 927.2341.

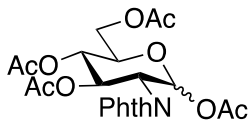
_p_-Nitrophenyl β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-α-D-galactopyranoside 3.13[16]



p-Nitrophenyl 2,3,4,6-tetra-*O*-benzoyl-β-D-galactopyranosyl-(1→3)-2-azido-2-deoxy-α-D-galactopyranoside **3.27** (130 mg, 0.17 mmol) was dissolved in dry pyridine (5 mL). Thioacetic acid (1.5 mL, 20.9 mmol) was added and the reaction was stirred at rt. After 36 h, t.l.c. (ethyl acetate/toluene, 1:1) indicated complete consumption of starting material (R_f 0.7) and the formation of a major product (R_f 0.1). The reaction mixture was diluted with DCM (50 mL) and the organic phase was washed with hydrochloric acid (100 mL of a 1.2 M solution), sodium hydrogen carbonate (100 mL of a saturation solution), brine (100 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. The residue was purified by flash column chromatography (ethyl acetate/toluene, 1:1) to give a white powder which was used directly for the next step. The residue was dissolved in dry methanol (10 mL) and methanolic sodium methoxide (10 mL of a 0.1 M solution) was added and the reaction mixture stirred overnight, at which point t.l.c.

(ethyl acetate/toluene, 1:1) indicated complete consumption of starting material (R_f 0.1) and formation of a single product (R_f 0). The reaction mixture was neutralized with Amberlite® IR120 (H^+), filtered, and concentrated *in vacuo*. The residue was purified by RP HPLC (column: Phenomenex Luna 5U C18 100 Å (250 x 10 mm x 10 µm); eluent: linear gradient of MeCN using a gradient method, 10-100% MeCN; flow rate: 3.5 $mLmin^{-1}$; column oven: 40 °C; detection: UV 210 nm) and concentrated *in vacuo* to give *p*-nitrophenyl β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-α-D-galactopyranoside **3.13** (6.4 mg, 9 %) as a white solid; HPLC: t_R = 7.86 min; m.p. 200 °C [lit. 212 – 214 °C];[18] $[\alpha]_D^{20}$ +11.3 (*c*, 0.5 in H_2O) [lit. $[\alpha]_D^{23}$ +10.2 (*c*, 0.5 in H_2O)];[18] δ_H (400 MHz, $CDCl_3$)[19] 1.92 (3H, s, CH_3CO_2), 3.47 (1H, at, *J* 7.8 Hz, H-2b), 3.55-3.85 (8H, m, H-5a, H-6a, H-6a', H-3b, H-4b, H-5b, H-6b, H-6b'), 3.91 (1H, m, H-4a), 4.20 (1H, m, H-3a), 4.44-4.51 (2H, m, H-1b, H-2a), 5.74 (1H, d, $J_{1a,2a}$ 3.6 Hz, H-1a), 7.19 (2H, d, *J* 9.4 Hz, $OC_6H_4NO_2$), 8.17 (2H, d, *J* 9 Hz, $OC_6H_4NO_2$); HRMS (ESI) Calcd. For $C_{20}H_{28}N_2O_{13}$ (MNa^+) 527.1489. Found 527.1488.

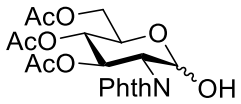
1,3,4,6-Tetra-*O*-acetyl-2-deoxy-2-phthalimido-D-glucopyranoside 3.30[20]



NaOH (2.4 g, 60.3 mmol) was added to a stirred solution of D-glucosamine hydrochloride **3.29** (10 g, 46.4 mmol) in methanol (20 mL) and water (40 mL). Phthalic anhydride (13.8 g, 92.8 mmol) was dissolved in acetone (80 mL) and added while maintaining the temperature below 15 °C. Sodium hydrogen carbonate (10.1 g, 121 mmol) was added, and the solution was subsequently stirred at rt. After 25 h, t.l.c. (ethyl acetate:acetic acid:methanol:water, 12:3:3:2) indicated the complete consumption of starting material (R_f 0.1) and the formation of a major product (R_f 0.3). The mixture was then brought to pH 1 with concentrated hydrochloric acid, concentrated *in vacuo*, and left at 2 °C in the refrigerator overnight. The resulting precipitate was filtered off, washed with cold water and dried under vacuum. The crude product was dissolved in pyridine (140 mL) and stirred at 0 °C. Acetic anhydride (105 mL, 1.1 mol) was then added and the mixture was stirred at rt overnight. At this point, t.l.c. (petrol:ethyl acetate, 1:1) indicated complete consumption of starting material (R_f 0) and the formation of a major compound (R_f 0.5). The reaction mixture was quenched by slowly adding it to ice-water (2 L). The organic phase was separated and washed with hydrochloric acid (6 x 1 L of a 1.2 M solution), sodium hydrogen carbonate (3 x 1 L of a saturated solution), brine (1 L), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. The crude product was purified by flash column chromatography (petrol:ethyl acetate, 2:1) to give 1,3,4,6-tetra-*O*-acetyl-2-deoxy-2-phthalimido-D-glucopyranoside **3.30** (16.6 g, 75 %) as a mixture of anomers (α/β, 2:1); δ_H (400 MHz, $CDCl_3$)[20] 1.84, 2.00,

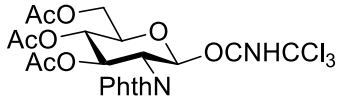
2.02, 2.08 (12H, 4 x s, 4 x CH_3CO_2), 3.97-4.02 (1H, m, H-5β), 4.04-4.09 (1H, m, H-6'α), 4.10-4.13 (1H, m, H-6β), 4.26-4.30 (1H, m, H-5α), 4.31-4.36 (2H, m, H-6α, H-6'β), 4.43 (1H, dd, $J_{1,2}$ 9.0 Hz, $J_{2,3}$ 10.6 Hz, H-2β), 4.68 (1H, dd, $J_{1,2}$ 3.2 Hz, $J_{2,3}$ 11.4Hz, H-2α), 5.10-5.30 (2H, m, H-4α, H-4β), 5.85 (1H, dd, $J_{3,4}$ 10.2 Hz, H-3β), 6.24 (1H, as, H-1α), 6.47 (H, as H-1β), 6.53 (1H, dd, $J_{3,4}$ 9.0 Hz, H-3α), 7.74-7.85 (4H, m, Phth); HRMS (ESI) Calcd. For $C_{22}H_{23}NO_{11}$ (MNa^+) 500.1163. Found 500.1167.

3,4,6-Tri-*O*-acetyl-2-deoxy-2-phthalimido-D-glucopyranoside 3.31



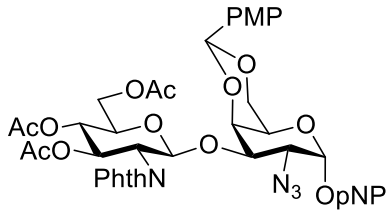
1,3,4,6-Tetra-*O*-acetyl-2-deoxy-2-phthalimido-D-glucopyranoside **3.30** (10 g, 21.0 mmol) was dissolved in dry DMF (100 mL). Hydrazine acetate (2.9 g, 31.4 mmol) was added and the reaction was stirred at rt. After 23 h, t.l.c. (petrol:ethyl acetate, 1:1) indicated complete consumption of starting material (R_f 0.5) and the formation of a major product (R_f 0.2). The reaction mixture was diluted with DCM (200 mL) and the organic phase was washed with brine (100 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. The residue was purified by flash column chromatography (petrol:ethyl acetate, 1:1) to give 3,4,6-tri-*O*-acetyl-2-deoxy-2-phthalimido-D-glucopyranoside **3.31** (6.3 g, 69 %) as a white solid, m.p. 170 °C [lit. m.p. 166-167 °C];[21] δ_H (400 MHz, $CDCl_3$)[22] (only α anomer quoted) 1.87, 2.04, 2.12 (9H, 3 x s, 3 x CH_3CO_2), 3.24 (1H, bd, $J_{1,OH}$ 6.7 Hz, 1-OH), 3.93 (1H, m, H-5), 4.18-4.29 (3H, m, H-2, H-6, H-6'), 5.17 (1H, at, *J* 9.4 Hz, H-4), 5.63 (1H, m, H-1), 5.85 (1H, at, *J* 9.8 Hz, H-3), 7.74 (2H, m, Phth), 7.85 (2H, m, Phth); HRMS (ESI) Calcd. For $C_{34}H_{28}O_{10}$ (MNa^+) 619.1575. Found 619.1599.

3,4,6-Tri-*O*-acetyl-2-deoxy-2-phthalimido-β-D-glucopyranosyl trichloroacetimidate 3.32



3,4,6-Tri-*O*-acetyl-2-deoxy-2-phthalimido-α-D-glucopyranoside **3.31** (300 mg, 0.69 mmol) and 1,1,1-trichloroacetonitrile (0.7 mL, 6.90 mmol) were stirred in dry DCM (10 mL) and cooled to 0 °C. DBU (20 μL, 0.14 mmol) was added. After 6 h, t.l.c. (petrol:ethyl acetate, 1:1) indicated complete consumption of starting material (R_f 0.2) and the formation of a major product (R_f 0.4). The reaction mixture was concentrated, and the residue purified by flash column chromatography (petrol:ethyl acetate, 1:1) to give 3,4,6-tri-*O*-acetyl-2-deoxy-2-phthalimido-β-D-glucopyranosyl trichloroacetimidate **3.32** (370 mg, 93 %) as a yellow oil; δ_H (400 MHz, $CDCl_3$)[15] 1.87, 2.03, 2.10 (9H, 3 x s, 3 x CH_3CO), 4.06 (1H, m, H-5), 4.18 (1H, m, H-6), 4.37 (1H, m, H-6'), 4.61 (1H, at, *J* 9.8 Hz, H-2), 5.26 (1H, at, *J* 9.3 Hz, H-4), 5.90 (1H, at, *J* 9.8 Hz, H-3), 6.61 (1H, d, $J_{1,2}$ 9 Hz, H-1), 7.72 (2H, m, Phth), 7.81 (2H, m, Phth), 8.65 (1H, s, N*H*); HRMS (ESI) Calcd. For $C_{22}H_{21}Cl_3N_2O_{10}$ (MNa^+) 601.0159. Found 601.0154.

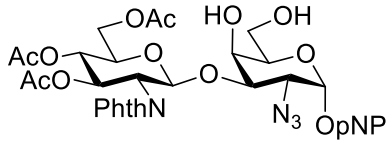
_p_-Nitrophenyl 3,4,6-tri-_O_-acetyl-2-deoxy-2-phthalimido-β-D-glucopyranosyl-(1→3)-2-azido-2-deoxy-4,6-_O_-_p_-methoxybenzylidene-α-D-galactopyranoside 3.33[16]



A mixture of *p*-nitrophenyl 2-azido-2-deoxy-4,6-*O*-*p*-methoxybenzylidene-α-D-galactopyranoside **3.22** (100 mg, 0.225 mmol), 3,4,6-tri-*O*-acetyl-2-deoxy-2-phthalimido-β-D-glucopyranosyl trichloroacetimidate **2.32** (300 mg, 0.527 mmol), and 4Å MS in dry DCM (40 mL) was stirred for 30 min under an N_2 atmosphere and then cooled to -20 °C. TMSOTf (4.1 μL, 0.023 mmol) was then added and the reaction mixture was stirred at -20 °C. After 7 h, t.l.c. (toluene:ethyl acetate, 2:1) indicated complete consumption of acceptor **3.22** (R_f 0.5) and the formation of a major product (R_f 0.4). Water (10 mL) was added. After 30 mins, the mixture was diluted with DCM (50 mL) and filtered through a pad of Celite®. The filtrate was washed with sodium hydrogen carbonate (50 mL of a saturated solution), brine (50 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. Purification by flash column chromatography (toluene:ethyl acetate, 4:1) afforded *p*-nitrophenyl 3,4,6-tri-*O*-acetyl-2-deoxy-2-phthalimido-β-D-glucopyranosyl-(1→3)-2-azido-2-deoxy-4,6-*O*-*p*-methoxybenzylidene-α-D-galactopyranoside **3.33** (140 mg, 72 %) as a white solid, m.p. 95-96 °C; $[\alpha]_D^{20}$ +97 (*c*, 1.0 in $CHCl_3$) [lit. $[\alpha]_D^{20}$ +135 (*c*, 0.57 in $CHCl_3$)][16] δ_H (400 MHz, $CDCl_3$)[16] 1.86, 2.04, (9H, 2 x s, 3 x CH_3CO_2), 3.67 (1H, m, H-5a), 3.80 (3H, s, $C_6H_3OCH_3$), 3.92-4.03 (3H, m, H-2a, H-5b, H-6a'), 4.16-4.24 (3H, m, H-3a, H-6a, H-

6b'), 4.42-4.53 (3H, m, H-4a, H-2b, H-6b), 5.22 (1H, at, *J* 9.4 Hz, H-4b), 5.53 (1H, s, C*H*$C_6H_4OCH_3$), 5.67 (1H, m, H-1a), 5.76-5.84 (2H, m, H-3b, H-1b), 6.87 (2H, m, $C_6H_4OCH_3$), 7.14 (2H, m, $OC_6H_4OCH_3$) 7.40 (2H, m, $C_6H_4OCH_3$) 7.70 (2H, m, Phth), 7.82 (2H, m, Phth), 8.16 (2H, m, $OC_6H_4NO_2$); HRMS (ESI) Calcd. For $C_{40}H_{39}N_5O_{17}$ (MNa^+) 884.2233. Found 884.2243.

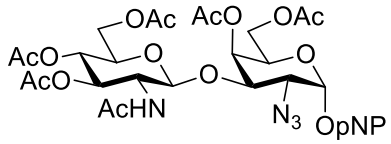
_p_-Nitrophenyl 3,4,6-tri-*O*-acetyl-2-deoxy-2-phthalimido-β-D-glucopyranosyl-(1→3)-2-azido-2-deoxy-α-D-galactopyranoside 3.34[16]



Acetic acid (80 % aq. solution, 10 mL) was added to a solution of *p*-nitrophenyl 3,4,6-tri-*O*-acetyl-2-deoxy-2-phthalimido-β-D-glucopyranosyl-(1→3)-2-azido-2-deoxy-4,6-*O*-*p*-methoxybenzylidene-α-D-galactopyranoside **3.33** (120 mg, 0.139 mmol) in DCM (10 mL) and stirred at rt. After 44 h, t.l.c. (toluene:ethyl acetate, 2:1) indicated complete consumption of starting material (R_f 0.4) and the formation of a major product (R_f 0.1). The mixture was diluted with DCM (50 mL) and carefully washed with sodium hydrogen carbonate (2 x 100 mL of saturated solution), brine (100 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. The crude product was purified by flash column chromatography (toluene:ethyl acetate, 1:1→0:1) to afford *p*-nitrophenyl 3,4,6-tri-*O*-acetyl-2-deoxy-2-phthalimido-β-D-glucopyranosyl-(1→3)-2-azido-2-deoxy-α-D-galactopyranoside **3.34** (70 mg, 68 %) as a white solid, m.p. 170 °C; $[\alpha]_D^{20}$ +169 (*c*, 1.0 in $CHCl_3$) [lit. $[\alpha]_D^{20}$ +213 (*c*, 0.63 in $CHCl_3$)];[16] δ_H (400 MHz, $CHCl_3$)[16] 1.89, 2.07, 2.13 (9H, 3 x s, 3 x CH_3CO_2), 3.76 (2H, m, H-6a, H-6a'), 3.85 (2H, m, H-2a, H-5a), 3.99 (1H, m, H-5b), 4.20 (2H, m, H-3a, H-6b'), 4.38 (2H, m, H-4a, H-6b), 4.45 (1H,

m, H-2b), 5.17 (1H, at, *J* 10.6 Hz, H-4b), 5.60 (1H, d, $J_{1,2}$ 3.4 Hz, H-1a), 5.75 (1H, d, $J_{1,2}$ 8.2 Hz, H-1b), 5.81 (1H, at, *J* 10.1 Hz, H-3b), 7.13 (2H, m, $OC_6H_4NO_2$), 7.75 (2H, m, Phth), 7.87 (2H, m, Phth), 8.18 (2H, m, $OC_6H_4NO_2$); HRMS (ESI) Calcd. For $C_{32}H_{33}N_5O_{16}$ (MNa^+) 766.1820. Found 766.1827.

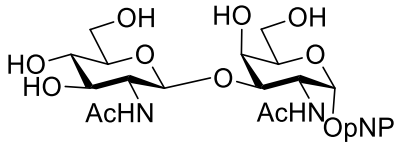
_p_-Nitrophenyl 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-β-D-glucopyranosyl-(1→3)-4,6-di-*O*-acetyl-2-azido-2-deoxy-α-D-galactopyranoside 3.35[16]



p-Nitrophenyl 3,4,6-tri-*O*-acetyl-2-deoxy-2-phthalimido-β-D-glucopyranosyl-(1→3)-2-azido-2-deoxy-α-D-galactopyranoside **3.34** (70 mg, 0.094 mmol) in dry methanol (10 mL) was added methanolic sodium methoxide (10 mL of a 0.1 M solution) and stirred overnight under N_2 atmosphere, at which point t.l.c. (toluene/ethyl acetate, 1:1) indicated complete consumption of starting material (R_f 0.4) and formation of a single product (R_f 0). The reaction mixture was neutralized with Amberlite® IR120 (H^+), filtered, and concentrated *in vacuo* to afford a white solid which was used directly for the next step. The crude residue was dissolved in ethanol (10 mL) and stirred at 70 °C. Hydrazine hydrate (70 μL, 1.41 mmol) was added. After 18 h, the reaction mixture was cooled to rt, concentrated *in vacuo*, and azeotroped with toluene to afford a yellow material which was used directly for the next step. The crude material was dissolved in pyridine (3 mL). Acetic anhydride (3 mL) and DMAP (2 mg, 0.02 mmol) were added and the reaction mixture was stirred overnight, at which point t.l.c. (ethyl acetate) indicated complete consumption of starting material (R_f 0) and formation of a major product (R_f 0.5). The reaction mixture was concentrated *in vacuo*, diluted with DCM

(50 mL), washed with hydrochloric acid (20 mL of a 1.2 M solution), sodium hydrogen carbonate (2 x 50 mL of a saturated solution), brine (50 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. Purification by flash column chromatography (toluene:ethyl acetate, 1:1) afforded *p*-nitrophenyl 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-β-D-glucopyranosyl-(1→3)-4,6-di-*O*-acetyl-2-azido-2-deoxy-α-D-galactopyranoside **3.35** (56 mg, 80 %) as a white solid, m.p. 180 - 185 °C; $[\alpha]_D^{20}$ +82 (*c*. 1.0 in $CHCl_3$) [lit. $[\alpha]_D^{20}$ +150 (*c*, 0.25 in $CHCl_3$)][16] δ_H (400 MHz, $CDCl_3$)[16] 1.91, 1.96, 2.04, 2.10, 2.16 (18H, 5 x s, 6 x CH_3CO_2), 3.72 (1H, m, H-5b), 3.84 (1H, m, H-2b), 3.95 (2H, m, H-2a, H-6a), 4.14 (3H, m, H-5a, H-6b, H-6a'), 4.30 (2H, m, H-3a, H-6b'), 4.99 (1H, d, $J_{1b,2b}$ 8.6 Hz, H-1b), 5.11 (1H, at, *J* 10.2 Hz, H-4b), 5.32 (1H, at, *J* 10.5 Hz, H-3b) 5.50 – 5.67 (2H, m, H-4a, N*H*), 5.66 (1H, d, $J_{1a,2a}$ 3.6 Hz, H-1a), 7.20 (2H, m, $OC_6H_4NO_2$), 8.23 (2H, m, $OC_6H_4NO_2$); HRMS (ESI) Calcd. For $C_{30}H_{37}N_5O_{17}$ (MNa^+) 762.2082. Found 762.2089.

_p_-Nitrophenyl 2-acetamido-2-deoxy-β-D-glucopyranosyl-(1→3)-2-acetamido-2-deoxy-α-D-galactopyranoside 3.14

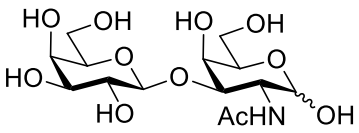


p-nitrophenyl 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-β-D-glucopyranosyl-(1→3)-4,6-di-*O*-acetyl-2-azido-2-deoxy-α-D-galactopyranoside **3.35** (160 mg, 0.216 mmol) was dissolved in dry pyridine (2 mL). Thioacetic acid (1.5 mL, 20.9 mmol) was added and the reaction was stirred at rt. After 49 h, t.l.c. (ethyl acetate/toluene, 1:1) indicated complete consumption of starting material (R_f 0.6) and the formation of a major product (R_f 0.1). The reaction mixture was diluted with DCM (50 mL) and the organic phase was washed with hydrochloric acid (100 mL of a 1.2 M solution), sodium hydrogen

carbonate (100 mL of a saturation solution), brine (100 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. The residue was purified by flash column chromatography (ethyl acetate→ethyl acetate/MeOH (9:1)) to give a white powder which was used directly for the next step. The residue was dissolved in dry methanol (10 mL) and methanolic sodium methoxide (10 mL of a 0.1 M solution) was added and the reaction mixture stirred for 1 h, at which point t.l.c. (ethyl acetate/toluene, 1:1) indicated complete consumption of starting material (R_f 0.1) and formation of a single product (R_f 0). The reaction mixture was neutralized with Amberlite® IR120 (H^+), filtered, and concentrated *in vacuo*. The residue was purified by RP HPLC (column: Phenomenex Luna 5U C18 100 Å (250 x 10 mm x 10 µm); eluent: linear gradient of MeCN using a gradient method, 10-100% MeCN; flow rate: 3.5 $mLmin^{-1}$; column oven: 40 °C; detection: UV 210 nm) and concentrated *in vacuo* to give *p*-nitrophenyl 2-acetamido-2-deoxy-β-D-glucopyranosyl-(1→3)-2-acetamido-2-deoxy-α-D-galactopyranoside

3.14 (10 mg, 11 %) as a white solid; HPLC: t_R = 9.63 min; m.p. 215-220 °C; $[\alpha]_D^{20}$ +237 (*c*, 0.6 in H_2O); δ_H (400 MHz, $CDCl_3$)[16] 1.92 (3H, s, CH_3CO_2), 2.11 (3H, s, CH_3CO_2), 3.35-3.39 (2H, m, H-4b, H-5b), 3.47 (1H, at, *J* 7.8 Hz, H-3b), 3.54-3.69 (4H, m, H-2b, H-6a, H-6a', H-6b), 3.78-3.82 (1H, m, H-6b'), 3.87 (1H, m, H-5a), 4.12 (1H, dd, $J_{2,3}$ 10.9 Hz, $J_{3,4}$ 2.7 Hz, H-3a), 4.39 (1H, dd, $J_{1,2}$ 3.9 Hz, H-2a), 4.55 (1H, d, $J_{1,2}$ 8.3 Hz, H-1b), 5.66 (1H, d, H-1a), 7.15 (2H, d, *J* 9.3 Hz, $OC_6H_4NO_2$), 8.15 (2H, d, *J* 9.3 Hz, $OC_6H_4NO_2$); HRMS (ESI) Calcd. For $C_{22}H_{31}N_3O_{13}$ (MNa^+) 568.1749. Found 568.1743.

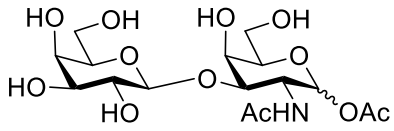
β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-D-galactopyranose 3.37



p-Nitrophenyl β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-D-galactopyranoside (10 mg, 37.99 μmol) was dissolved in sodium acetate buffer (100 μL of a 25 mM, pH 6.0 solution). EngEF (2 μL of a 1.5 mg/mL solution) was added and the reaction mixture was incubated at 37 ºC. The reaction mixture was analysed by HPLC and UV until the complete disappearance of starting material after 1 h. The product was isolated by RP-HPLC (column: Phenomenex Luna 5U C18 100 Å (250 x 10 mm x 10 μm); eluent A: water, eluent B: MeCN; 0-9 min: 100% A, 9-12 min 100% B, 12-15 min: 100 % A, 15-20 min: 10% A; flow rate: 2.0 $mLmin^{-1}$; column oven: 15 °C; detection: CAD) to give β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-D-galactopyranose **3.37** (15 mg, 99 %) as a white solid; t_R = 7.22 min (α anomer), 8.08 min (β anomer); ν_{max} (neat) 3298 (s, OH), 1625 (s, C=O) cm^{-1}; δ_H (400 MHz, D_2O) 1.89 (6H, s, 2 x CO_2CH_3), 3.36-3.41 (2H, m, H-2b$_\alpha$, H-2b$_\beta$), 3.46-3.54 (10H, m, H-6a$_\alpha$, H-6'a$_\alpha$, H-6a$_\beta$, H-6'a$_\beta$, H-6b$_\alpha$, H-6b'$_\beta$,H-6b$_\beta$, H-6'$_\beta$), 3.57 (6H, m, H-3b$_\alpha$, H-3b$_\beta$, H-5a$_\alpha$, H-5a$_\beta$, H-5b$_\alpha$, H-5b$_\beta$), 3.72 (1H, dd, $J_{2,3}$ 11 Hz, $J_{3,4}$ 3.6 Hz, H-3a$_\alpha$), 3.78 (1H, m, H-4b$_\beta$), 3.85 (1H, m, H-3a$_\beta$), 3.89 (1H, dd, $J_{1,2}$ 3.9 Hz, $J_{2,3}$ 10.9 Hz, H-2a$_\alpha$), 4.00 (1H, at, *J* 6.3 Hz, H-4b$_\alpha$), 4.04 (1H, d, *J* 3.1 Hz, H-4a$_\beta$), 4.11 (1H, d, *J* 3.1 Hz, H-4a$_\alpha$), 4.16 (1H, dd, $J_{1,2}$ 3.9 Hz, H-2a$_\alpha$), 4.30 (1H, d, $J_{1,2}$ 7.9 Hz, H-1b$_\alpha$), 4.36 (1H, d, $J_{1,2}$ 7.9 Hz, H-1b$_\beta$), 4.56 (1H, d, $J_{1,2}$ 9.0 Hz, H-1a$_\beta$), 5.07 (1H, d, $J_{1,2}$ 3.6 Hz, H-1a$_\alpha$); δ_C (100 MHz, D_2O) 21.9, 22.2 (q, 2 x CH_3), 48.9 (d, C-2a$_\alpha$), 52.4 (d C-2a$_\beta$), 60.9, 60.9, 61.1, 62.7 (4 x t, 4 x C-6), 68.0 (d, C-3b$_\alpha$), 68.5 (d, C-4a$_\beta$), 68.7 (d, C-4a$_\alpha$), 70.1 (d, C-5a$_\beta$), 70.5 (d, C-3a$_\alpha$),

50.6 (d, C-2b$_\alpha$), 72.4 (d, C-3b$_\beta$), 72.5 (d, C-2b$_\beta$), 74.8 (d, C-5b$_\alpha$), 74.9 (d, C-5b$_\beta$), 77.0 (d, C-4b$_\alpha$), 80.0 (d, C-4b$_\beta$), 91.1 (d, C-1a$_\alpha$), 95.1 (d, C-1a$_\beta$), 104.6 (d, C-1b$_\alpha$), 104.8 (d, C-1b$_\beta$), 174.6, 174.9 (2 x s, 2 x C=O); HRMS (ESI) Calcd. For $C_{14}H_{25}NO_{11}$ (MNa^+) 406.1325. Found 406.1315.

1-*O*-Acetyl-β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-D-galactopyranose



β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-D-galactopyranose **3.37**

(5 mg, 13.0 μmol), triethylamine (16.3 μL, 117 μmol), and thioacetic acid (4.7 μL, 65.2 μmol) were stirred in D_2O (52.2 μL) and cooled to 0 °C. 2-Chloro-1,3-dimethylimidazolinium chloride (6.64 μg, 39.1 μmol) dissolved in MeCN (10 μL) was added in 2 μL portions. After 30 mins, the reaction mixture was diluted with water (400 μL), washed with DCM (2 x 400 μL), filtered through a column of Amberlite® IR120 (H^+) previously activated by 1M NaOH solution, and then lyophilized. The product was purified by semi-preparative RP (column: Phenomenex Luna 5U C18 100 Å (250 x 10 mm x 10 μm); eluent A: water, eluent B: MeCN; 0-30 min: 100% A, 30-40 min 100% B, 40-60 min: 100 % A, 60-70 min: 100% A; flow rate: 1.0 $mLmin^{-1}$; column oven: 15 °C; detection: CAD) to afford 1-*O*-acetyl-β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-D-galactopyranose (4.8 mg, 87 %) as a white foam (α/β, 1:1.1), t_R (α anomer) = 42.89 mins; t_R (β anomer) = 46.35 mins; ν_{max} (neat) 3304 (s, OH), 2929 (s, CH), 1738 (s, C=O), 1593 (s, N-H) cm^{-1}; δ_H (400 MHz, D_2O) (α anomer) 1.86 (3H, s, NHCOC$\underline{H}_3$), 2.06 (3H, m, $COCH_3$), 3.42 (1H, m, H-2b), 3.49-3.65 (7H, m, H-3b, H-5a, H-5b, H-6a,

H-6a', H-6b, H-6b'), 3.78 (1H, m, H-4b), 3.97 (1H, m, H-4a), 4.17 (1H, m, H-3a), 4.33-4.36 (2H, m, H-1b, H-2a), 5.96 (1H, d, $J_{1a,2a}$ 3.5 Hz, H-1a); (β anomer) 1.87 (3H, s, NHCOC$\underline{H}_3$), 2.00 (3H, s, $COCH_3$), 3.37-3.49 (2H, m, H-2b, H-6a), 3.50-3.54 (2H, m, H-3b, H-6b), 3.61-3.71 (4H, m, H-5a, H-5b, H-6a', H-6b'), 3.78 (1H, m, H-4b), 3.87 (1H, dd, $J_{2,3}$ 11.0 Hz, $J_{3,4}$ 3.0 Hz, H-3a), 4.03-4.10 (2H, m, H-2a, H-4a), 4.33 (1H, d, $J_{1b,2b}$ 7.4 Hz, H-1b), 5.50 (1H, d, $J_{1a,2a}$ 9.0 Hz, H-1a); HRMS (ESI) Calcd. For $C_{16}H_{27}NO_{12}$ (MNa^+) 448.1425. Found 448.1425.

Experimental for Chapter 4

Standard kinetic assay

The standard assay for GH101 endo-α-*N*-acetylgalactosaminidase activity monitored the release of p-nitrophenol by gain of absorbance at 235 nm. All kinetic measurements were made using a Varian Cary 100 UV-visible spectrophotometer, stoppered 1 cm pathlength quartz cuvettes and a total assay volume of 1 mL. Standard assays were carried out at 37 °C and all assay solutions were equilibrated to temperature and baseline absorbance before initialisation. The enzymatic reaction was initiated by the addition of 2 μL of purified enzyme at a stock concentration of 2-3 mg/mL. Standard reaction mixtures contained 25 mM NaOAc buffered at pH 7.5, and a variable concentration of Galβ1,3GalNAc-α-pNP. Metal-ion solutions were created using Milli-Q water. Initial rates of reaction were measured as a least-squares fit of the initial rate data using Cary WinUV Kinetics Application (version 3.00, Varian). The values of the kinetic parameters were determined by fitting data to the Michaelis-Menten equation using the software GraFit 5 (Erithacus Software Limited).

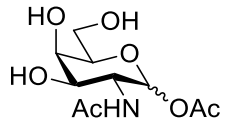
Kinetic assay for glycosyl acetates

Glycosyl acetates lack a chromophore which was problematic for determining the kinetic parameters for hydrolysis. The kinetic parameters could not be determined using UV spectroscopy. Therefore, the hydrolysis reaction was monitored using a charged aerosol detector, whereby initial rates of hydrolysis could be inferred by integrating the peak areas, calculating hydrolysis yield, and dividing by the time of the reaction to obtain a rate. The kinetic parameters were then obtained by fitting these experimental data to the Michaelis-Menten model using the non-linear regression program, GraFit 5 (Erithacus Software).

General endo-α-*N*-acetylgalactosaminidase-catalysed reaction procedure

Activated glycoside (1 equiv) and acceptor (10 equiv) were dissolved in 25 mM NaOAc buffer (pH 6.0) and warmed to 37 ˚C. Enzyme was added and the reaction mixture was incubated for 16 h before quenching with 1 M NaOH. The reaction mixture was then analysed by RP-HPLC (column: Phenomenex Luna 5U C18 100 Å (250 x 10 mm x 10 µm); eluent A: water, eluent B: MeCN; 0-30 min: 100% A, 30-40 min 100% B, 40-60 min: 100 % A, 60-70 min: 100% A; flow rate: 1.0 $mLmin^{-1}$; column oven: 15 °C; detection: CAD/UV 210 nm).

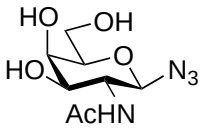
Acetyl 2-Acetamido-2-deoxy-D-galactopyranoside 4.1 & 4.2



N-Acetyl-D-galactosamine (20 mg, 0.09 mmol), triethylamine (125 µL, 0.9 mmol), and thioacetic acid (32 µL, 0.452 mmol) were stirred in D_2O (500 µL) and cooled to 0 °C. 2-Chloro-1,3-dimethylimidazolinium chloride (46 mg, 0.27 mmol) was added. After 30 mins, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.2) and the formation of a major product (R_f 0.5). The reaction mixture was diluted with water (10 mL), washed with DCM (5 x 20 mL), lyophilised to 1 mL, filtered through a column of Amberlite® IR120 (H^+) previously activated by 1M NaOH solution, and then lyophilized. The product was purified by semi-preparative RP HPLC (column: Luna 5U C18 (100 Å) column (Phenomenex); eluent: H_2O only. Sample was

run at 1 mL/min for 40 mins; column oven: 40 °C; detection: CAD) to afford gave acetyl 2-acetamido-2-deoxy-D-galactopyranoside **4.1** & **4.2** (18 mg, 86 %) as a white foam (α/β, 1:1.1), t_R = 11.65 mins; ν_{max} (neat) 3273 (s, OH), 1742 (s, C=O), 1642 (s, N-H) cm^{-1}; δ_H (400 MHz, D_2O) 1.91 (6H, s, α-NHCOCH_3, β-NHCOCH_3), 2.04 (3H, s, β-COCH_3), 2.08 (3H, s, α-CO<u>C</u>H_3), 3.63-3.68 (7H, H-4β, H-5α, H-5β, H-6α, H-6'α, H-6β, H-6'β), 3.89-3.97 (4H, m, H-2α, H-3α, H-3β, H-4α), 4.20 (1H, m, H-2α), 5.49 (1H, d, $J_{1,2}$ 8.6 Hz, H-1β), 6.01 (1H, s, H-1α); δ_C (100 MHz, D_2O) 20.2, 20.2, 21.7, 22.0 (4 x q, 4 x CH_3CO_2), 48.7 (d, C-2α), 51.3 (d, C-2β), 60.7, 61.0 (2 x t, C-6α. C-6β), 67.2, 67.5 (2 x d, C-3α, C-3β), 68.2, 70.4 (2 x d, C-4α, C-4β), 73.2, 76.1 (2 x d, C-5α, C-5β), 91.2 (d, C-1α), 93.3 (d, C-1β), 172.4, 172.8, 174.8, 175.0 (4 x s, 4 x <u>C</u>OCH_3); HRMS (ESI) Calcd. For $C_{10}H_{17}NO_7$ (MNa^+) 286.0903. Found 286.0904.

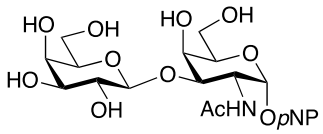
2-Acetamido-2-deoxy-β-D-galactopyranosyl azide 4.3



N-Acetyl-D-galactosamine (5 g, 22.6 mmol) and triethylamine (15.6 mL, 113.0 mmol) were stirred in H_2O/MeCN (4:1, 100 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate (19.3 g, 67.8 mmol) was added. After 3 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.2) and the formation of two major products (R_f 0.5, 0.6). The reaction mixture was acidified to pH 2 by dropwise addition of aqueous 1.2M HCl and then neutralized by

addition of sat. $NaHCO_3$ solution (100 mL). The reaction mixture was concentrated, and the residue dissolved in ethanol (20 mL), filtered through Celite®, and concentrated *in vacuo*. The residue was then redissolved in water (50 mL), washed with DCM (2 x 50 mL), filtered through a column of Amberlite® IR120 (H^+, previously treated with 1M NaOH solution), and concentrated *in vacuo*. Purification by flash column chromatography ($CHCl_3$:MeOH, 7:1 → $CHCl_3$:MeOH, 2:1) gave 2-acetamido-2-deoxy-β-D-galactopyranosyl azide **4.3** (4.3 g, 77 %) as a white solid; m.p. 195-200 °C; $[\alpha]_D^{20}$ -16 (*c*, 1.0 in MeOH); ν_{max} (neat) 1643 (s, amide), 2094 (s, azide), 3331.3 (bs, OH) cm^{-1}; δ_H (400 MHz, D_2O) 1.94 (3H, s, CH_3), 3.63-3.78 (4H, m, H-3, H-5, H-6, H-6'), 3.80 (1H, at, *J* 10.2 Hz, H-2), 3.86 (1H, ad, *J* 3.1 Hz, H-4), 4.55 (1H, d, $J_{1,2}$ 9.4 Hz, H-1); δ_C (100 MHz, D_2O) 22.1 (d, NHCO<u>C</u>H_3), 51.7 (d, C-2), 60.9 (t, C-6), 67.6 (d, C-4), 70.7 (d, C-3), 77.2 (d, C-5), 89.0 (d, C-1), 175.0 (s, NH<u>C</u>OCH_3) HRMS (ESI) Calcd. For $C_8H_{14}N_4O_5$ (MNa^+) 269.0862. Found 269.0854.

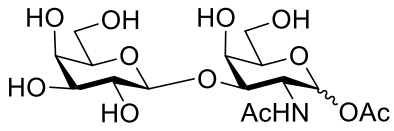
_p_-Nitrophenyl β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-α-D-galactopyranoside 4.4



p-Nitrophenyl 2,3,4,6-tetra-*O*-benzoyl-β-D-galactopyranosyl-(1→3)-2-azido-2-deoxy-α-D-galactopyranoside (130 mg, 0.17 mmol) was dissolved in dry pyridine (5 mL). Thioacetic acid (1.5 mL, 20.9 mmol) was added and the reaction was stirred at rt. After

36 h, t.l.c. (ethyl acetate/toluene, 1:1) indicated complete consumption of starting material (R_f 0.7) and the formation of a major product (R_f 0.1). The reaction mixture was diluted with DCM (50 mL) and the organic phase was washed with hydrochloric acid (100 mL of a 1.2 M solution), sodium hydrogen carbonate (100 mL of a saturation solution), brine (100 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. The residue was purified by flash column chromatography (ethyl acetate/toluene, 1:1) to give a white powder which was used directly for the next step. The residue was dissolved in dry methanol (10 mL) and methanolic sodium methoxide (10 mL of a 0.1 M solution) was added and the reaction mixture stirred overnight, at which point t.l.c. (ethyl acetate/toluene, 1:1) indicated complete consumption of starting material (R_f 0.1) and formation of a single product (R_f 0). The reaction mixture was neutralized with Amberlite® IR120 (H^+), filtered, and concentrated *in vacuo*. The residue was purified by RP HPLC (column: Phenomenex Luna 5U C18 100 Å (250 x 10 mm x 10 µm); eluent: linear gradient of MeCN using a gradient method, 10-100% MeCN; flow rate: 3.5 $mLmin^{-1}$; column oven: 40 °C; detection: UV 210 nm) and concentrated *in vacuo* to give *p*-nitrophenyl β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-α-D-galactopyranoside **4.4** (6.4 mg, 9 %) as a white solid; HPLC: t_R = 7.86 min; m.p. 200 °C [lit. 212 – 214 °C]; $[\alpha]_D^{20}$ +11.3 (*c*, 0.5 in H_2O) [lit. $[\alpha]_D^{23}$ +10.2 (*c*, 0.5 in H_2O)]; δ_H (400 MHz, $CDCl_3$) 1.92 (3H, s, CH_3CO_2), 3.47 (1H, at, *J* 7.8 Hz, H-2b), 3.55-3.85 (8H, m, H-5a, H-6a, H-6a', H-3b, H-4b, H-5b, H-6b, H-6b'), 3.91 (1H, m, H-4a), 4.20 (1H, m, H-3a), 4.44-4.51 (2H, m, H-1b, H-2a), 5.74 (1H, d, $J_{1a,2a}$ 3.6 Hz, H-1a), 7.19 (2H, d, *J* 9.4 Hz, $OC_6H_4NO_2$), 8.17 (2H, d, *J* 9 Hz, $OC_6H_4NO_2$); HRMS (ESI) Calcd. For $C_{20}H_{28}N_2O_{13}$ (MNa^+) 527.1489. Found 527.1488.

1-*O*-Acetyl-β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-D-galactopyranose 4.7 & 4.8

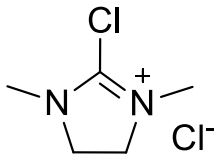


β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-D-galactopyranose (5 mg, 13.0 μmol), triethylamine (16.3 μL, 117 μmol), and thioacetic acid (4.7 μL, 65.2 μmol) were stirred in D_2O (52.2 μL) and cooled to 0 °C. 2-Chloro-1,3-dimethylimidazolinium chloride (6.64 μg, 39.1 μmol) dissolved in MeCN (10 μL) was added in 2 μL portions. After 30 mins, the reaction mixture was diluted with water (400 μL), washed with DCM (2 x 400 μL), filtered through a column of Amberlite® IR120 (H^+) previously activated by 1M NaOH solution, and then lyophilized. The product was purified by semi-preparative RP (column: Phenomenex Luna 5U C18 100 Å (250 x 10 mm x 10 μm); eluent A: water, eluent B: MeCN; 0-30 min: 100% A, 30-40 min 100% B, 40-60 min: 100 % A, 60-70 min: 100% A; flow rate: 1.0 $mLmin^{-1}$; column oven: 15 °C; detection: CAD) to afford gave 1-*O*-acetyl-β-D-galactopyranosyl-(1→3)-2-acetamido-2-deoxy-D-galactopyranose (4.8 mg, 87 %) as a white foam (α/β, 1:1.1), t_R (α anomer **4.7**) = 42.89 mins; t_R (β anomer **4.8**) = 46.35 mins; ν_{max} (neat) 3304 (s, OH), 2929 (s, CH), 1738 (s, C=O), 1593 (s, N-H) cm^{-1}; δ_H (400 MHz, D_2O) (α anomer) 1.86 (3H, s, NHCOC$\underline{H_3}$), 2.06 (3H, m, $COCH_3$), 3.42 (1H, m, H-2b), 3.49-3.65 (7H, m, H-3b, H-5a, H-5b, H-6a, H-6a', H-6b, H-6b'), 3.78 (1H, m, H-4b), 3.97 (1H, m, H-4a), 4.17 (1H, m, H-3a), 4.33-4.36 (2H, m, H-1b, H-2a), 5.96 (1H, d, $J_{1a,2a}$ 3.5 Hz, H-1a); (β anomer) 1.87 (3H, s, NHCOC$\underline{H_3}$), 2.00 (3H, s, $COCH_3$), 3.37-3.49 (2H, m, H-2b, H-6a), 3.50-

3.54 (2H, m, H-3b, H-6b), 3.61-3.71 (4H, m, H-5a, H-5b, H-6a', H-6b'), 3.78 (1H, m, H-4b), 3.87 (1H, dd, $J_{2,3}$ 11.0 Hz, $J_{3,4}$ 3.0 Hz, H-3a), 4.03-4.10 (2H, m, H-2a, H-4a), 4.33 (1H, d, $J_{1b,2b}$ 7.4 Hz, H-1b), 5.50 (1H, d, $J_{1a,2a}$ 9.0 Hz, H-1a); HRMS (ESI) Calcd. For $C_{16}H_{27}NO_{12}$ (MNa^+) 448.1425. Found 448.1425.

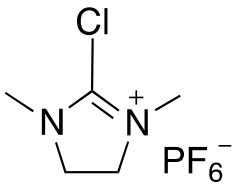
Experimental for Chapter 5

2-Chloro-1,3-dimethylimidazolinium chloride 5.1[23]



1,3-Dimethylimidazolidinone (18.9 mL, 175 mmol) was dissolved in toluene (200 mL) and stirred under an N_2 atmosphere. Oxalyl chloride (19.8 mL, 219 mmol) was added dropwise and the reaction was stirred at 50 °C. After 3 h, the flask was allowed to cool to rt and stirred overnight. After 18 h, the precipitate was filtered, washed with cold toluene and recrystallized (acetonitrile/diethyl ether). The white solid was filtered and dried under vacuum to yield 2-chloro-1,3-dimethylimidazolidinium chloride **5.1** (18.1 g, 61 %) as a white powdery solid, m.p. 83-85 °C [lit. m.p. 95-100 °C];[24] δ_H (400 MHz, CD_3CN)[25] 3.17 (6H, s, 2 x CH_3), 4.02 (4H, s, 2 x CH_2); δ_C (100 MHz, CD_3CN)[25] 34.4, 50.0; HRMS (ESI) Calcd. For $C_5H_{10}ClN_2$ (M^+) 133.0527. Found 133.0529.

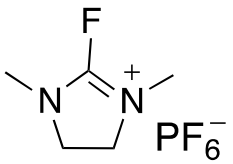
2-Chloro-1,3-dimethylimidazolinium hexafluorophosphate 5.2[26]



2-Chloro-1,3-dimethylimidazolidinium chloride **5.1** (4 g, 23.7 mmol) was dissolved in anhydrous acetonitrile (50 mL) and stirred under an N_2 atmosphere. $NaPF_6$ (4 g, 23.7

mmol) was added and the reaction stirred at rt. After 30 mins, the solution was filtered through Celite® and washed with acetonitrile (30 mL). The filtrate was concentrated *in vacuo* and the resultant solid was dissolved in a minimal amount of acetonitrile. Et_2O was added until a precipitate formed, which was collected by suction filtration to afford 2-chloro-1,3-dimethylimidazolinium hexafluorophosphate **5.2** (6.60 g, 99 %) as a white solid, m.p. 210 °C [lit. 230-231 °C];[26] δ_H (400 MHz, CD_3CN)[26] 3.14 (6H, s, 2 x CH_3), 3.95 (4H, s, 2 x CH_2); δ_C (100 MHz, CD_3CN)[26] 34.2, 49.8; δ_F (376.23 MHz, CD_3CN) -72.08 (d, J = 705.7 Hz, PF_6); δ_P (161.86 MHz, CD_3CN) -143.36 (sep, J = 706.6 Hz, PF_6); HRMS (ESI) Calcd. For $C_5H_{10}ClN_2$ (M^+) 133.0527. Found 133.0529.

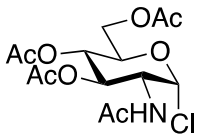
2-Fluoro-1,3-dimethylimidazolinium hexafluorophosphate 5.3[27]



2-Chloro-1,3-dimethylimidazolidinium hexafluorophosphate **5.2** (6.60 g, 23.8 mmol) was dissolved in anhydrous acetonitrile (60 mL) and stirred at 60 °C under an N_2 atmosphere. KF (4.14 g, 71.2 mmol), which had been dried with a heat gun under vacuum, was added to the reaction mixture. After 4 h, the mixture was cooled to rt and filtered through Celite® washed with acetonitrile. The filtrate was concentrated *in vacuo* and the resulting solid dissolved in a minimal amount of acetonitrile. Et_2O was added until a precipitate formed, which was collected by suction filtration to afford 2-fluoro-1,3-dimethylimidazolinium hexafluorophosphate **5.3** (3.76 g, 60 %) as a white solid, m.p. 150-155 °C [lit. 168-169 °C];[27] δ_H (400 MHz, CD_3CN)[27] 3.88 (d, 4H, J = 2.4 Hz,

CH_2), 3.02 (s, 6H, CH_3); δ_C (100 MHz, CD_3CN)[27] 29.9 (s, CH_3), 46.7 (s, CH_2); δ_P (161.86 MHz, CD_3CN) -143.4 (sep, *J* = 705.7 Hz, *P*F_6); δ_F (376.23 MHz, CD_3CN) -73.0 (d, *J* = 705.7 Hz, P*F*$_6$), -90.5 (s, C-*F*); HRMS (ESI) Calcd. For $C_5H_{10}FN_2$ (M^+) 117.0823. Found 117.0825.

2-Acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-α-D-glucopyranosyl chloride 5.7[28]

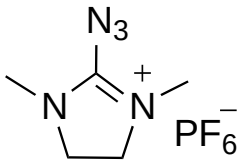


N-Acetyl-D-glucosamine **5.4** (1 g, 4.52 mmol) was suspended in acetyl chloride (100 mL). After 22 h, t.l.c. (petrol:ethyl acetate, 1:2) indicated complete consumption of starting material (R_f 0) and the formation of two major products (R_f 0.2, 0.4). The reaction mixture was diluted with DCM (100 mL) and the organic phase was washed with ice-water (100 mL), sodium hydrogen carbonate (2 x 100 mL of a saturated solution), brine (100 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. The residue was purified by flash column chromatography (ethyl acetate) to give 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-α-D-glucopyranosyl chloride **5.7** (410 mg, 25 %) as a yellow oil, $[\alpha]_D^{20}$ +118 (*c*, 1.0 in $CHCl_3$); δ_H (400 MHz, $CDCl_3$)[28] 1.98, 2.06, 2.10 (12H, 3 x s, 4 x CH_3CO), 4.15 (1H, m, H-6), 4.30, (2H, m, H-5, H-6'), 4.53 (1H, m, H-2), 5.21 (1H, at, *J* 10.5 Hz, H-4), 5.31 (1H, at, *J* 9.8 Hz, H-3), 5.75 (1H, d, $J_{2,NH}$ 7 Hz, NH), 6.19 (1H, m, H-1); HRMS (ESI) Calcd. For $C_{14}H_{20}ClNO_8$ (MNa^+) 388.0775. Found 388.0769.

2-Acetamido-2-deoxy-β-D-glucopyranosyl azide 5.9

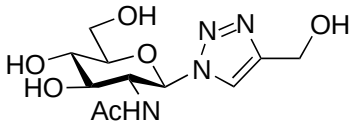
N-Acetyl-D-glucosamine **5.4** (500 mg, 2.26 mmol) and triethylamine (1.6 mL, 11.3 mmol) were stirred in D_2O/MeCN (4:1, 10 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (1.9 g, 6.78 mmol) was added. After 1 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.2) and the formation of two major products (R_f 0.5, 0.6). The reaction mixture was acidified to pH 2 by dropwise addition of aqueous 1.2 M HCl and then neutralized by addition of sat. $NaHCO_3$ solution (10 mL). The reaction mixture was concentrated, and the residue dissolved in ethanol (20 mL), filtered through Celite®, and concentrated *in vacuo*. The residue was then redissolved in water (10 mL), washed with DCM (2 x 20 mL), filtered through a column of Amberlite® IR120 (H^+, previously treated with 1 M NaOH solution), and concentrated *in vacuo*. Purification by flash column chromatography ($CHCl_3$:MeOH, 5:1) gave 2-acetamido-2-deoxy-β-D-glucopyranosyl azide **5.9** (480 mg, 86 %) as a white solid; m.p. 125 - 130 °C; $[\alpha]_D^{20}$ -34 (*c*, 1.0 in MeOH); δ_H (400 MHz, D_2O)[29] 1.88 (3H, s, CH_3), 3.32 (1H, m, H-4), 3.35-3.42 (2H, m, H-3, H-5), 3.53 (1H, at, *J* 9.5 Hz, H-2), 3.60 (1H, dd, $J_{5,6}$ 5.4 Hz, $J_{6,6'}$ 12.2 Hz, H-6), 3.75 (1H, m, H-6'), 4.58 (1H, d, $J_{1,2}$ 9.3 Hz, H-1); HRMS (ESI) Calcd. For $C_8H_{14}N_4O_5$ (MNa^+) 269.0862. Found 269.0854.

2-Azido-1,3-dimethylimidazolinium hexafluorophosphate 5.10[26]



2-Chloro-1,3-dimethylimidazolidinium hexafluorophosphate **5.2** (11.6 g, 41.64 mmol) was dissolved in anhydrous acetonitrile (120 mL) and stirred at 0 °C under an N_2 atmosphere. NaN_3 (3.8 g, 58.5 mmol) was added to the reaction mixture. After 3 h, the mixture was filtered through Celite® washed with acetonitrile. The filtrate was concentrated *in vacuo* and the resulting solid dissolved in a minimal amount of acetonitrile. Et_2O was added until a precipitate formed, which was collected by suction filtration to afford 2-azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (11.8 g, quant.) as a white solid, m.p. 200-205 °C [lit. 205-207 °C];[26] δ_H (400 MHz, CD_3CN)[26] 3.08 (s, 6H, CH_3), 3.80 (s, 4H, CH_2); δ_P (161.86 MHz, CD_3CN); δ_F (376.23 MHz, CD_3CN); HRMS (ESI) Calcd. For $C_5H_{10}N_5$ (M^+) 140.0931. Found 140.0925.

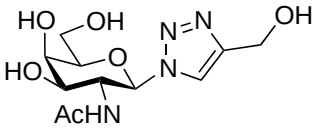
1-(2'-Acetamido-2'-deoxy-β-D-glucopyranosyl)-4-hydroxymethyl-1,2,3-triazole 5.12



N-Acetyl-D-glucosamine **5.4** (500 mg, 2.26 mmol) and triethylamine (1.6 mL, 11.3 mmol) were stirred in D_2O/MeCN (4:1, 10 mL) and cooled to 0 °C. 2-Azido-1,3-

dimethylimidazolinium hexafluorophosphate **5.10** (1.9 g, 6.78 mmol) was added. After 1 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.2) and the formation of two major products (R_f 0.5, 0.6). The reaction mixture was acidified to pH 2 by dropwise addition of aqueous 1.2 M HCl and then neutralized by addition of sat. aqueous $NaHCO_3$ (10 mL). Propargyl alcohol (0.3 mL, 4.52 mmol), $CuSO_4.5H_2O$ (8.5 mg, 0.03 mmol), and L-ascorbic acid (80 mg, 0.45 mmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.5) and formation of a single product (R_f 0.2). The reaction mixture was concentrated, and the residue was dissolved in ethanol (20 mL), filtered through Celite®, and concentrated *in vacuo*. The residue was then redissolved in water, washed with DCM (2 x 20 mL), filtered through a column of Amberlite® IR120 (H^+, previously treated by 1 M NaOH solution), and concentrated *in vacuo*. Purification by flash column chromatography ($CHCl_3$:MeOH, 3:1) gave 1-(2'-acetamido-2'-deoxy-β-D-glucopyranosyl)-4-hydroxymethyl-1,2,3-triazole **5.12** (580 mg, 85 %) as a white solid; m.p. 80-85 °C ($CHCl_3$/MeOH); $[\alpha]_D^{20}$ -99 (*c*, 1.0 in MeOH); ν_{max} (neat) 3268 (s, OH), 1650 (s, C=O), 1548 (s, NH) cm^{-1}; δ_H (400 MHz, D_2O) 1.67 (3H, s, NHCOC<u>H</u>$_3$), 3.52-3.70 (4H, m, H-3, H-4, H-5, H-6), 3.78 (1H, m, *J* 12.5 Hz, H-6'), 4.11 (1H, t, *J* 10 Hz, H-2), 4.56 (2H, s, CH_2), 5.69 (1H, d, $J_{1,2}$ 9.8 Hz, H-1), 8.03 (1H, s, C=CH); δ_C (100 MHz, D_2O) 21.5 (q, C(O)<u>C</u>H$_3$), 54.4 (t, C<u>C</u>H$_2$OH), 55.3 (d, C-2), 60.3 (t, C-6), 69.2 (d, C-4), 73.4 (d, C-3), 78.8 (d, C-5), 86.3 (d, C-1), 122.6 (s, <u>C</u>H=C), 147.0 (s, CH=<u>C</u>), 174.1 (s, NH<u>C</u>OCH$_3$); HRMS (ESI) Calcd. For $C_{11}H_{18}N_4O_6$ (MNa^+) 325.1124. Found 325.1117.

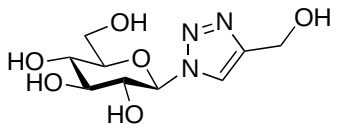
1-(2'-Acetamido-2'-deoxy-β-D-galactopyranosyl)-4-hydroxymethyl-1,2,3-triazole 5.14



N-Acetyl-D-galactosamine **5.13** (500 mg, 2.26 mmol) and triethylamine (1.6 mL, 11.3 mmol) were stirred in D_2O/MeCN (4:1, 10 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (1.9 g, 6.78 mmol) was added. After 1 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.2) and the formation of two major products (R_f 0.5, 0.6). The reaction mixture was acidified to pH 2 by dropwise addition of 1.2 M HCl solution and then neutralized by addition of sat. $NaHCO_3$ solution (10 mL). Propargyl alcohol (0.3 mL, 4.52 mmol), $CuSO_4.5H_2O$ (8.5 mg, 0.03 mmol), and L-ascorbic acid (80 mg, 0.45 mmol) were added and the reaction mixture was stirred at 50 °C overnight, at which time, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.5) and formation of a single product (R_f 0.2). The reaction mixture was concentrated, and the residue was dissolved in ethanol (20 mL), filtered through Celite®, and concentrated *in vacuo*. The residue was then redissolved in water (10 mL), washed with DCM (2 x 20 mL), filtered through a column of Amberlite® IR120 (H^+) previously activated by 1 M NaOH solution, and concentrated *in vacuo*. Purification by flash column chromatography ($CHCl_3$:MeOH, 3:1) gave 1-(2'-acetamido-2'-deoxy-β-D-galactopyranosyl)-4-hydroxymethyl-1,2,3-triazole **5.14** (510 mg, 75 %) as a white solid; m.p. 95-100 °C; $[\alpha]_D^{20}$ +26 (*c*, 1.0 in MeOH); ν_{max} (neat) 3284 (s, OH), 1648 (s, C=O), 1547 (s, NH) cm^{-1}; δ_H (500 MHz, D_2O) 1.66 (3H, s, NHCOC$\underline{H_3}$), 3.68 (2H, m,

H-6, H-6'), 3.83 (2H, m, H-3, H-5), 3.95 (1H, d, *J* 1.2 Hz, H-4), 4.28 (1H, t, J 10.6 Hz, H-2), 4.57 (2H, s, CH_2), 5.62 (1H, d, $J_{1,2}$ 9.5 Hz, H-1), 8.07 (1H, s, C=CH); δ_C (125 MHz, D_2O) 21.6 (q, NHCO<u>C</u>H_3), 51.9 (d, C-2), 54.4 (t, <u>C</u>H_2), 60.8 (t, C-6), 67.6 (d, C-4), 70.5 (d, C-5), 78.2 (d, C-3), 86.8 (d, C-1), 122.5 (d, <u>C</u>H=C), 147.0 (s, CH=<u>C</u>), 174.3 (s, NH<u>C</u>OCH_3); HRMS (ESI) Calcd. For $C_{11}H_{18}N_4O_6$ (MNa^+) 325.1124. Found 325.1117.

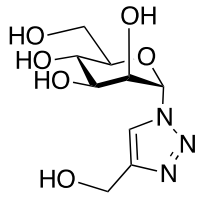
1-(β-D-Glucopyranosyl)-4-hydroxymethyl-1,2,3-triazole 5.16



D-Glucose **5.15** (500 mg, 2.78 mmol) and triethylamine (1.9 mL, 13.9 mmol) were stirred in D_2O/MeCN (4:1, 10 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (2.4 g, 8.33 mmol) was added. After 3 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.3) and the formation of a major product (R_f 0.5). Propargyl alcohol (0.32 mL, 5.55 mmol), $CuSO_4.5H_2O$ (10 mg, 0.04 mmol), and L-ascorbic acid (100 mg, 0.56 mmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.5) and formation of a single product (R_f 0.2). The reaction mixture was concentrated, and the residue was dissolved in ethanol (20 mL), filtered through Celite®, and concentrated *in vacuo*. The residue was then redissolved in water (10 mL), washed with DCM (2 x 20 mL), filtered through a column of Amberlite® IR120 (H^+, previously treated with 1 M NaOH solution), and concentrated *in vacuo*. Purification by flash column

chromatography ($CHCl_3$:MeOH, 3:1) gave 1-(β-D-glucopyranosyl)-4-hydroxymethyl-1,2,3-triazole **5.16** (530 mg, 73 %) as a white solid; m.p. 65-70 °C; $[\alpha]_D^{20}$ -4 (*c*, 1.0 in MeOH); ν_{max} (neat) 3274 (s, OH) cm^{-1}; δ_H (500 MHz, D_2O) 3.48 (1H, at, *J* 9.3 Hz, H-4), 3.55-3.66 (3H, m, H-3, H-5, H-6), 3.77 (1H, m, H-6'), 3.87 (1H, t, *J* 9.3 Hz, H-2), 4.62 (2H, s, CH_2), 5.61 (1H, d, $J_{1,2}$ 9.1 Hz, H-1), 8.07 (1H, s, C=CH); δ_C (125 MHz, D_2O) 54.5 (t, CH_2), 60.3 (t, C-6), 68.8 (d, C-4), 72.1 (d, C-2), 75.8 (d, C-3), 78.8 (d, C-5), 87.3 (d, C-1), 123.2 (d, CH=C), 147.0 (s, CH=C); HRMS (ESI) Calcd. For $C_9H_{15}N_3O_6$ (MNa^+) 284.0859. Found 284.0852.

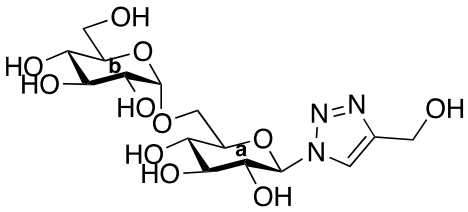
1-(α-D-Mannopyranosyl)-4-hydroxymethyl-1,2,3-triazole 5.18



D-Mannose **5.17** (500 mg, 2.78 mmol) and triethylamine (1.9 mL, 13.9 mmol) were stirred in D_2O/MeCN (4:1, 10 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (2.4 g, 8.33 mmol) was added. After 3 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.3) and the formation of a major product (R_f 0.5). Propargyl alcohol (0.32 mL, 5.55 mmol), $CuSO_4.5H_2O$ (10 mg, 0.04 mmol), and L-ascorbic acid (100 mg, 0.56 mmol) were added and the reaction mixture was stirred at 50 °C overnight, at which time, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.5) and formation of a single product (R_f 0.2). The reaction mixture was concentrated, and the residue was dissolved in ethanol (20 mL), filtered through Celite®, and concentrated *in*

vacuo. The residue was then redissolved in water (10 mL), washed with DCM (2 x 20 mL), filtered through a column of Amberlite® IR120 (H^+, previously treated with 1 M NaOH solution), and concentrated *in vacuo*. Purification by flash column chromatography ($CHCl_3$:MeOH, 3:1) gave 1-(α-D-mannopyranosyl)-4-hydroxymethyl-1,2,3-triazole **5.18** (530 mg, 73 %) as a white solid; m.p. 45-50 °C; $[\alpha]_D^{20}$ +43 (*c*, 1.0 in MeOH); ν_{max} (neat) 3276 (s, OH) cm^{-1}; δ_H (500 MHz, D_2O) 3.18 (1H, m, H-5), 3.56 (1H, m, H-4), 3.53 (1H, m, H-6), 3.58 (1H, m, H-6'), 4.01 (1H, dd, $J_{2,3}$ 3.4 Hz, $J_{3,4}$ 9.0 Hz, H-3), 4.61 (2H, s, CH_2), 4.65 (1H, m, H-2), 5.99 (1H, d, $J_{1,2}$ 0.7 Hz, H-1), 8.01 (1H, s, C=CH); δ_C (125 MHz, D_2O) 54.4 (t, CH_2), 60.4 (t, C-6), 66.4 (d, C-4), 68.2 (d, C-2), 70.4 (d, C-3), 76.0 (d, C-5), 86.5 (d, C-1), 123.6 (d, CH=C), 147.1 (s, CH=C); HRMS (ESI) Calcd. For $C_9H_{15}N_3O_6$ (MNa^+) 284.0859. Found 284.0847.

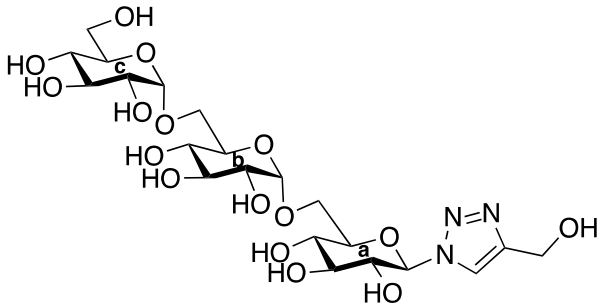
1-(6-*O*-α-D-Glucopyranosyl-β-D-glucopyranosyl)-4-hydroxymethyl-1,2,3-triazole 5.20



Isomaltose **5.19** (250 mg, 0.73 mmol) and triethylamine (0.5 mL, 3.7 mmol) were stirred in D_2O/MeCN (4:1, 5 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (630 mg, 2.2 mmol) was added. After 3 h, propargyl alcohol (85 μL, 1.5 mmol), $CuSO_4.5H_2O$ (3 mg, 0.011 mmol), and L-ascorbic acid (30 mg, 0.15 mmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, the reaction mixture was concentrated, and the residue

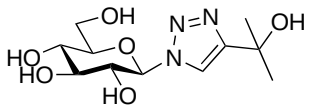
was dissolved in methanol (20 mL), filtered through Celite®, and concentrated *in vacuo*. The residue was then redissolved in water, washed with DCM (2 x 20 mL), filtered through a column of Amberlite® IR120 (H^+, previously treated with 1 M NaOH solution), and concentrated *in vacuo*. Purification by flash column chromatography ($CHCl_3$:MeOH, 3:1 → H_2O:*i*PrOH:EtOAc, 1:6:3) gave 1-(6-*O*-α-D-glucopyranosyl-β-D-glucopyranosyl)-4-hydroxymethyl-1,2,3-triazole **5.20** (267 mg, 86 %) as a white solid; m.p. 95-100 °C; $[\alpha]_D^{20}$ +61 (*c*, 1.0 in MeOH); ν_{max} (neat) 3295 (s, OH), 1016 (s, O-C-O) cm^{-1}; δ_H (500 MHz, D_2O) 3.24 (1H, at, *J* 9.3 Hz, H-4b), 3.37 (1H, dd, $J_{1b,2b}$ 3.6 Hz, $J_{2b,3b}$ 10.0 Hz, H-2b), 3.49 (3H, m, H-3b, H-6a, H-6'a), 3.54 (1H, m, H-5b), 3.56 (2H, m, H-3a, H-4a), 3.72 (1H, m, H-6b), 3.81 (1H, m, H-6'b), 3.92 (1H, at, *J* 9.0 Hz, H-2a) , 4.61 (2H, s, CH_2), 4.76 (1H, d, H-1b), 5.62 (1H, d, $J_{1a,2a}$ 9.3 Hz, H-1a), 8.08 (1H, s, C=CH); δ_C (125 MHz, D_2O) 54.4 (t, CH_2), 60.0 (t, C-6a), 65.9 (t, C-6b), 68.9 (d, C-4a), 69.1 (d, C-4b), 71.2 (d, C-2b), 71.6 (d, C-3b), 71.9 (d, C-2a), 72.9 (d, C-5b), 76.0 (d, C-3a), 77.4 (d, C-5a), 87.2 (d, C-1a), 98.0 (d, C-1b), 123.4 (d, CH=C), 146.9 (s, CH=C); HRMS (ESI) Calcd. For $C_{15}H_{24}DN_3O_{11}$ (MNa^+) 447.1450. Found 447.1455.

1-[α-D-Glucosyl-[1→6]-α-D-glucosyl-[1→6]-β-D-glucopyranosyl)-4-hydroxymethyl-1,2,3-triazole 5.22



Isomaltotriose **5.21** (125 mg, 0.25 mmol) and triethylamine (0.17 mL, 1.24 mmol) were stirred in D_2O/MeCN (4:1, 2 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (210 mg, 0.74 mmol) was added. After 3 h, propargyl alcohol (27 μL, 0.50 mmol), $CuSO_4.5H_2O$ (1 mg, 0.0037 mmol), and L-ascorbic acid (9 mg, 0.050 mmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, the reaction mixture was concentrated. The residue was dissolved in methanol (20 mL), filtered through Celite®, and concentrated *in vacuo*. The residue was then redissolved in water (10 mL), washed with DCM (2 x 20 mL), filtered through a column of Amberlite® IR120 (H^+, previously treated with 1 M NaOH solution), and concentrated *in vacuo*. The reaction mixture was purified by gel filtration on a column (8.7 x 1.7 cm) of Sephadex G-25 (pre-equilibrated and eluted with 0.01% ammonia) to give 1-[α-D-glucosyl-[1→6]-α-D-glucosyl-[1→6]-β-D-glucopyranosyl)-4-hydroxymethyl-1,2,3-triazole **5.22** (128 mg, 88 %) as a white solid; m.p. 45-50 °C; $[\alpha]_D^{20}$ +34 (*c*, 0.45 in H_2O); ν_{max} (neat) 3273 (s, OH), 1008 (s, O-C-O) cm^{-1}; δ_H (400 MHz, D_2O) 3.27-3.43 (6H, m, H-2b, H-4b, H-6b, H-2c, H-3c, H-4c), 3.53-3.57 (4H, m, H-4a, H-3b, H-5b, H-5c), 3.60-3.77 (4H, m, H-3a, H-6a, H-6'a, H-6'b), 3.80-3.83 (3H, m, H-5a, H-6c, H-6'c), 3.95 (1H, at, *J* 9.0 Hz, H-2a), 4.66 (2H, s, CH_2), 4.73 (1H, d, $J_{1b,2b}$ 3.5 Hz, H-1b), 4.80 (1H, d, $J_{1c,2c}$ 3.9 Hz, H-1c), 5.65 (1H, d, $J_{1a,2a}$ 9.0 Hz, H-1a); 8.12 (1H, s, C=CH); δ_C (100 MHz, D_2O) 54.6 (t, CH_2), 60.3 (t, C-6a), 64.9 (t, C-6b), 66.3 (t, C-6c), 69.1 (d, C-4c), 69.1 (d, C-4a) 69.4 (d, C-4b), 70.1 (d, C-2c), 71.2 (d, C-2b), 71.4 (d, C-3c), 71.7 (d, C-3b), 71.9 (d, C-2a), 73.0 (d, C-5b), 73.2 (d, C-5c), 76.0 (d, C-3a), 77.4 (d, C-5a), 87.3 (d, C-1a), 97.6 (d, C-1c), 98.0 (d, C-1b), 123.8 (d, CH=C), 146.0 (s, CH=C); HRMS (ESI) Calcd. For $C_{21}H_{36}N_3O_{16}$ (MH^+) 586.2096. Found 586.2093.

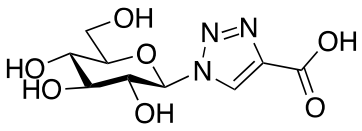
1-(β-D-Glucopyranosyl)-4-(dimethylhydroxy)methyl-1,2,3-triazole 5.23



D-Glucose **5.15** (500 mg, 2.78 mmol) and triethylamine (1.9 mL, 13.9 mmol) were stirred in D_2O/MeCN (4:1, 10 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (2.4 g, 8.33 mmol) was added. After 1 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.3) and the formation of a major product (R_f 0.5). 2-Methyl-3-butyn-2-ol (0.54 mL, 5.55 mmol), $CuSO_4.5H_2O$ (10 mg, 0.04 mmol), and L-ascorbic acid (100 mg, 0.56 mmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.5) and the formation of a single product (R_f 0.3). The reaction mixture was concentrated, and the residue was dissolved in ethanol (20 mL), filtered through Celite®, and concentrated *in vacuo*. The residue was then redissolved in water (10 mL), washed with DCM (2 x 20 mL), filtered through a column of Amberlite® IR120 (H^+, previously treated with 1 M NaOH solution), and concentrated *in vacuo*. Purification by flash column chromatography ($CHCl_3$:MeOH, 5:1 → $CHCl_3$:MeOH, 2:1) gave 1-(β-D-glucopyranosyl)-4-(dimethylhydroxy)methyl-1,2,3-triazole **5.23** (650 mg, 81 %) as a white solid; m.p. 45-50 °C; $[\alpha]_D^{20}$ +1 (*c*, 1.0 in MeOH); ν_{max} (neat) 3288 (bs, OH) cm^{-1}; δ_H (500 MHz, D_2O) 1.47 (6H, s, 2 x CH_3), 3.47 (1H, at, *J* 9.6 Hz, H-4), 3.57 (1H, m, H-3), 3.61 (1H, m, H-5), 3.63 (1H, m, H-6), 3.76 (1H, dd, $J_{5,6}$ 1.8 Hz, $J_{6,6'}$ 12.3 Hz, H-6'), 3.85 (1H, at, *J* 9.3 Hz, H-2), 5.58 (1H, d, $J_{1,2}$ 9.3 Hz, H-1), 8.01 (1H, s, C=CH); δ_C (125 MHz, D_2O) 28.7 (q, 2 x CH_3), 60.3 (t, C-6), 68.1 (d, C-4), 68.8 (s, $\underline{C}(CH_3)_2OH$),

72.1 (d, C-2), 75.8 (d, C-3), 78.7 (d, C-5), 87.3 (d, C-1), 120.8 (d, CH=C), 154.9 (s, CH=C); HRMS (ESI) Calcd. For $C_{11}H_{19}N_3O_6$ (MH^+) 290.1352. Found 290.1338.

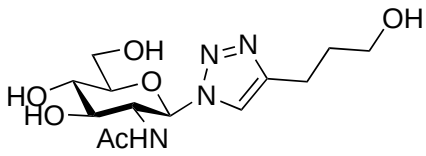
1-(β-D-Glucopyranosyl)-4-carboxy-1,2,3-triazole 5.24



D-Glucose **5.15** (500 mg, 2.78 mmol) and triethylamine (1.9 mL, 13.9 mmol) were stirred in D_2O/MeCN (4:1, 10 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (2.4 g, 8.33 mmol) was added. After 3 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.3) and the formation of a major product (R_f 0.5). Propiolic acid (0.34 mL, 5.55 mmol), $CuSO_4.5H_2O$ (10 mg, 0.04 mmol), and L-ascorbic acid (100 mg, 0.56 mmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.5) and formation of a single product (R_f 0). The reaction mixture was concentrated, and the residue was dissolved in ethanol (20 mL), filtered through Celite®, and concentrated *in vacuo*. The residue was then redissolved in water (10 mL), washed with DCM (2 x 20 mL), filtered through a column of Amberlite® IR120 (H^+, previously treated with 1 M NaOH solution), and concentrated *in vacuo*. Purification by reverse phase chromatography (H_2O → H_2O/MeOH, 9:1) gave 1-(β-D-glucopyranosyl)-4-carboxy-1,2,3-triazole **5.24** (740 mg, 97 %) as a yellow solid; m.p. 45-50 °C; $[\alpha]_D^{20}$ -2 (*c*, 1.0 in H_2O); ν_{max} (neat) 3284 (bs, OH) cm^{-1}; δ_H (400 MHz, D_2O) 3.62 (1H, t, *J* 3.9 Hz, H-4), 3.56 (1H, m, H-6), 3.70 (1H, m, H-3), 3.72 (1H, m, H-5), 3.99 (1H, t, *J* 9.3 Hz, H-2),

4.60 (2H, s, CH_2), 5.71 (1H, d, $J_{1,2}$ 9.1 Hz, H-1), 8.40 (1H, s, C=CH); δ_C (125 MHz, D_2O) 60.5 (t, C-6), 69.0 (d, C-4), 72.3 (d, C-2), 75.9 (d, C-5), 78.9 (d, C-3), 87.5 (d, C-1), 126.6 (d, C=<u>C</u>H), 145.0 (s, <u>C</u>=CH), 167.3 (s, C=O); HRMS (ESI) Calcd. For $C_9H_{12}DN_3O_7$ (MNa^+) 299.0714. Found 299.0701.

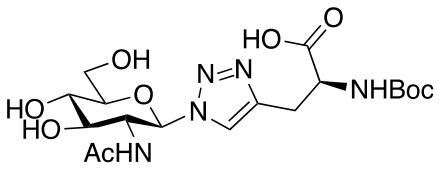
1-(2'-Acetamido-2'-deoxy-β-D-glucopyranosyl)-4-hydroxypropyl-1,2,3-triazole 5.25



N-Acetyl-D-glucosamine **5.4** (500 mg, 2.26 mmol) and triethylamine (1.6 mL, 11.3 mmol) were stirred in D_2O/MeCN (4:1, 10 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (1.9 g, 6.78 mmol) was added. After 1 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.2) and the formation of two major products (R_f 0.5, 0.6). The reaction mixture was acidified to pH 2 by dropwise addition of 1.2 M aqeuous HCl and then neutralized by addition of sat. aqueous $NaHCO_3$ (10 mL). 4-Pentyn-1-ol (0.42 mL, 4.52 mmol), $CuSO_4.5H_2O$ (8.5 mg, 0.03 mmol), and L-ascorbic acid (80 mg, 0.45 mmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.5) and formation of a single product (R_f 0.3). The reaction mixture was concentrated, and the residue was dissolved in ethanol (20 mL), filtered through Celite®, and concentrated *in vacuo*. The residue was then redissolved in water (10 mL), washed with DCM (2 x 20 mL), filtered through a column of Amberlite® IR120 (H^+, previously treated with 1 M NaOH solution), and concentrated *in vacuo*. Purification by flash column

chromatography ($CHCl_3$:MeOH, 3:1) gave 1-(2'-acetamido-2'-deoxy-β-D-glucopyranosyl)-4-hydroxypropyl-1,2,3-triazole **5.25** (710 mg, 95 %) as a white solid; m.p. 55-60 °C; $[\alpha]_D^{20}$ -5 (*c*, 1.0 in MeOH); ν_{max} (neat) 3261 (s, OH), 1648 (s, C=O), 1548 (s, N-H) cm^{-1}; δ_H (500 MHz, D_2O) 1.65 (3H, s, $NHCOCH_3$), 1.72 (2H, t, *J* 7.1 Hz, CH_2), 2.61 (2H, t, *J* 7.3 Hz, CH_2), 3.42 (2H, t, *J* 6.4 Hz, CH_2), 3.51-3.70 (4H, m, H-3, H-4, H-5, H-6), 3.77 (1H, m, H-6'), 4.07 (1H, t, *J* 9.8 Hz, H-2), 5.63 (1H, d, $J_{1,2}$ 9.6 Hz, H-1), 7.83 (1H, s, C=CH); δ_C (125 MHz, D_2O) 20.8 (q, $NHCO\underline{C}H_3$), 21.5 (t, CH_2), 30.8 (t, CH_2), 55.3 (d, C-2), 60.3 (t, C-6), 60.5 (t, CH_2), 69.2 (d, C-4), 73.4 (d, C-3), 78.8 (d, C-5), 86.2 (d, C-1), 121.7 (d, $\underline{C}H$=C), 147.9 (s, CH=$\underline{C}$), 173.9 (s, C=O); HRMS (ESI) Calcd. For $C_{13}H_{22}N_4O_6$ (MNa^+) 353.1437. Found 353.1423.

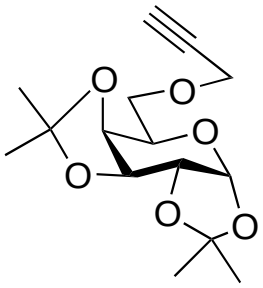
1-(2'-Acetamido-2'-deoxy-β-D-glucopyranosyl)-4-((S)-2-(Boc-amino)-4-pentanoic acid)-1,2,3-triazole 5.26



N-Acetyl-D-glucosamine **5.4** (130 mg, 0.59 mmol) and triethylamine (0.4 mL, 2.94 mmol) were stirred in D_2O/MeCN (4:1, 2 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (500 mg, 1.76 mmol) was added. After 1 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.2) and the formation of two major products (R_f 0.5, 0.6). The reaction mixture was acidified to pH 2 by dropwise addition of 1.2 M aqueous HCl and then neutralized by addition of sat. aqueous $NaHCO_3$ solution (10 mL). (*S*)-2-(Boc-amino)-4-pentynoic acid (250 mg, 1.2 mmol), $CuSO_4.5H_2O$ (2.2 mg, 8.8 μmol), and L-ascorbic

acid (21 mg, 0.12 mmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, the reaction mixture was filtered through Celite®, and concentrated *in vacuo*. The residue was then redissolved in water, washed with DCM (2 x 20 mL), filtered through a column of Amberlite® IR120 (H^+, previously activated by 1 M NaOH solution), and concentrated *in vacuo*. Purification by reverse phase chromatography (H_2O → H_2O/MeOH, 9:1) gave 1-(2'-acetamido-2'-deoxy-β-D-glucopyranosyl)-4-((S)-2-(Boc-amino)-4-pentanoic acid)-1,2,3-triazole **5.26** (210 mg, 78 %) as a white solid; m.p. 200-210 °C; $[\alpha]_D^{20}$ +15 (*c*, 1.0 in MeOH); ν_{max} (neat) 3275 (s, OH), 1709 (s, C=O) cm^{-1}; δ_H (400 MHz, D_2O) 1.22 (9H, s, Boc), 1.67 (3H, s, $NHCOCH_3$), 2.85-3.11 (2H, m, $C_\alpha HC\underline{H}_2$), 3.52-3.69 (4H, m, H-3, H-4, H-5, H-6), 3.77 (1H, d, H-6'), 4.10 (2H, m, H-2, $C_\alpha\underline{H}CH_2$), 5.65 (1H, d, $J_{1,2}$ 9.3 Hz, H-1), 7.84 (1H, s, C=CH); δ_C (100 MHz, D_2O) 21.6 (q, $NHCOCH_3$), 27.6 (t, $CH\underline{C}H_2$), 28.1 (q, $C(\underline{C}H_3)_3$), 55.3 (d, C-2), 60.4 (t, C-6), 69.2 (d, C-4), 73.6 (d, C-3), 78.8 (d, C-5), 81.1 (s, $\underline{C}(CH_3)_3$), 86.1 (d, C-1), 122.8 (d, $\underline{C}$H=C), 144.2, 157.2, 174.0 (3 x s, 3 x C=O); HRMS (ESI) Calcd. For $C_{18}H_{29}N_5O_9$ (MNa^+) 482.1863. Found 482.1856.

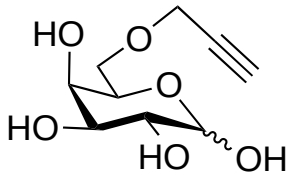
1,2:3,4-Di-*O*-isopropylidine-6-*O*-prop-2-yn-1-yl-α-D-galactopyranose 5.29[30]



Diacetone galactose **5.28** (1.5 g, 7.56 mmol) and KOH (740 mg, 13.3 mmol) were stirred in dry acetonitrile (10 mL) at rt under an N_2 atmosphere. Propargyl bromide (1

mL, 13.3 mmol) was added dropwise and the reaction mixture was stirred 16 h, after which time, t.l.c (petrol/ethyl acetate, 3:1) indicated the complete consumption of starting material (R_f 0.1) and the formation of a major product (R_f 0.4). The reaction mixture was concentrated *in vacuo.* The residue was dissolved in DCM (20 mL), washed with brine (30 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. Purification by flash column chromatography (petrol/ethyl acetate, 4:1) gave 1,2:3,4-di-*O*-isopropylidine-6-*O*-prop-2-yn-1-yl-α-D-galactopyranose **5.29 (**1.3 g, 76 %) as a white solid; m.p. 50-55 °C; $[\alpha]_D^{20}$ -83 (*c*, 1.0, $CHCl_3$); δ_H (400 MHz, $CDCl_3$)[30] 1.33, 1.34, 1.45, 1.54 (12H, 4 x s, 4 x CH_3), 2.42 (1H, t, *J* 2.4 Hz, $H_2CCC\underline{H}$, 3.67 (1H, dd, $J_{5,6}$ 7 Hz, $J_{6,6'}$ 10.1 Hz, H-6), 3.77 (1H, dd, $J_{5,6'}$ 5.4 Hz, H-6'), 4.01 (1H m, H-5), 4.22 (2H, dd, *J* 2.4 Hz, 8.2 Hz, $\underline{H_2}CCCH$), 4.26 (1H, dd, $J_{3,4}$ 8.2 Hz, $J_{4,5}$ 2.3 Hz, H-4), 4.31 (1H. dd, $J_{1,2}$ 5.1 Hz, $J_{2,3}$ 2.4 Hz, H-2), 4.60 (1H, dd, $J_{2,3}$ 2.4 Hz, $J_{3,4}$ 7.9 Hz, H-3), 5.54 (1H, d, H-1); HRMS (ESI) Calcd. For $C_{15}H_{22}O_6$ (MNa^+) 321.1314. Found 321.1306.

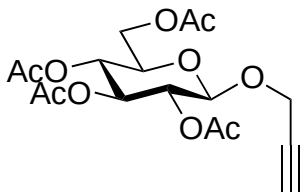
6-*O*-Prop-2-yn-yl-D-galactopyranose 5.30[30]



1,2:3,4-Di-*O*-isopropylidine-6-*O*-prop-2-yn-1-yl-α-D-galactopyranose **5.29** (4.8 g, 16.1 mmol) was stirred in water (100 mL). Dowex® 50WX8 (H^+) (6 g) was added and the reaction mixture was heated at 80 °C and stirred overnight. The reaction mixture was then filtered, and lyophilized to give 6-*O*-prop-2-yn-yl-D-galactopyranose **5.30** (3.1 mg, 88 %) as a colourless oil; δ_H (500 MHz, CD_3OD) 2.78 (2H, m, 2 x $CH_2CC\underline{H}$),

3.38 (1H, m, H-2β), 3.50-3.86 (11H, m, H-2α, H-3α, H-3β, H-4α, H-4β, H-5α, H-5β, H-6α, H-6β, H-6'α, H-6'β), 4.14 (4H, m, 2 x C$\underline{H_2}$CCH), 4.45 (1H, d, $J_{1,2}$ 7.8 Hz, H-1β), 5.12 (1H, d, $J_{1,2}$ 3.5 Hz, H-1α); HRMS (ESI) Calcd. For $C_9H_{14}O_6$ (MNa^+) 241.0688. Found 241.0680.

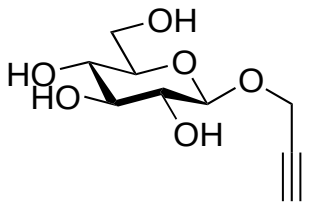
2'-Propynyl 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranoside 5.32[31]



Penta-*O*-acetyl-β-D-glucopyranose (16 g, 41.00 mmol) and propargyl alcohol (4.8 mL, 82.00 mmol) were stirred in dry DCM (400 mL) under an N_2 atmosphere and the solution was cooled to 0 °C. Boron trifluoride diethyl etherate (6.1 mL, 49.2 mmol) was added, and the mixture was allowed to warm to room temperature. After 24 h, the reaction mixture was diluted with DCM (200 mL), washed with sodium hydrogen carbonate (2 x 500 mL of a saturated aqueous solution), brine (500 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo.* The residue was recrystallized (ethanol) to afford 2'-propynyl 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranoside **5.32** (12.5 g, 79 %) as a white crystalline solid; m.p. 100-105 °C [lit. 116-117 °C];[31] $[\alpha]_D^{20}$ -49 (*c,* 1.0 in $CHCl_3$) [lit. $[\alpha]_D^{23}$ –43.4 (*c*, 0.9 in $CHCl_3$)];[31] δ_H (400 MHz, $CDCl_3$)[31] 2.00, 2.03, 2.06, 2.09 (12H, 4 x s, 4 x CH_3CO_2), 2.47 (1H, t, *J* 2.3 Hz, CCH), 3.73 (1H, ddd, $J_{4,5}$ 7.0 Hz, $J_{5,6}$ 2.3 Hz, $J_{5,6'}$ 4.7 Hz, H-5), 4.15 (1H, dd, $J_{6,6'}$ 12.1 Hz, H-6), 4.27 (1H, dd, H-6'), 4.37 (2H, d, *J* 2.3 Hz, CH_2), 4.78 (1H, d, $J_{1,2}$ 8.2 Hz, H-1), 5.01 (1H, t, H-2), 5.10 (1H, t, H-4), 5.25

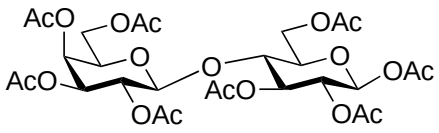
(1H, t, *J* 9.3 Hz, H-3); HRMS (ESI) Calcd. For $C_{17}H_{22}O_{10}$ (MNa^+) 409.1111. Found 409.1103.

2'-Propynyl-β-D-glucopyranoside 5.33[32]



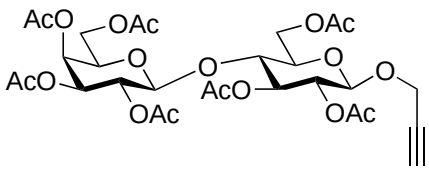
2'-Propynyl 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranoside **5.32** (10 g, 25.9 mmol) was stirred in dry methanol (200 mL). Sodium metal (440 mg) was added portionwise. After 1 h, t.l.c. (petrol:ethyl acetate, 1:1) indicated complete consumption of starting material (R_f 0.7) and the formation of a single product (R_f 0). The reaction mixture was neutralized with Dowex® 50WX8 (H^+), filtered, and concentrated *in vacuo* to give 2'-propynyl-β-D-glucopyranoside **5.33** (5.4 g, 96 %) as a yellow solid; m.p. 85-90 °C; $[\alpha]_D^{20}$ -97 (*c*, 1.0 in CH_3OH) [lit. $[\alpha]_D^{23}$ –91.9 (*c*, 0.66 in CH_3OH)];[32] δ_H (400 MHz, D_2O)[32] 2.79 (1H, t, *J* 2.3 Hz, CH_2CH), 3.17 (1H, t, *J* 9.0 Hz, H-2), 3.26 (1H, m, H-4), 3.32-3.41 (2H, m, H-3, H-5), 3.60 (1H, dd, $J_{5,6}$ 5.9 Hz, $J_{6,6'}$ 12.5 Hz, H-6), 3.80 (1H, dd, $J_{5,6'}$ 2.0 Hz, $J_{6,6'}$ 12.5 Hz, H-6'), 4.34 (2H, t, *J* 2.3 Hz, CH_2CH), 4.51 (1H, d, $J_{1,2}$ 8.2 Hz, H-1); HRMS (ESI) Calcd. For $C_9H_{14}O_6$ (MNa^+) 241.0688. Found 241.0679.

1,2,3,6,2',3',4',6'-Octa-*O*-acetyl-β-D-lactoside 5.35[33]



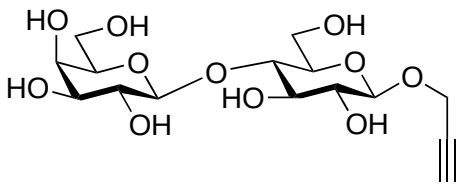
Sodium acetate (13.2 g, 161 mmol) was added to acetic anhydride (160 mL) and refluxed at 120 °C for 1 h. D-Lactose **5.34** (100 g, 292 mmol) was then added portion-wise while the reaction mixture was vigorously stirred. After the final portion was added, the solution became clear and was allowed to stir for an additional 1 h. The reaction mixture was allowed to cool to rt, and then poured into ice-cold water (5 L). The product was then extracted with DCM (3 x 1 L), washed with sodium hydrogen carbonate (6 x 500 mL of a saturated solution), brine (2 x 500 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo*. Recrystallization (ethanol) afforded 1,2,3,6,2',3',4',6'-Octa-*O*-acetyl-β-D-lactoside **5.35** (135 g, 68 %) as a white solid; m.p. 125-130 °C [lit. m.p. 140-143 °C];[33] $[\alpha]_D^{20}$ -2 (*c*, 1.0 in $CHCl_3$) [lit. $[\alpha]_D^{27}$ -4.0 (*c*, 3.60 in $CHCl_3$)];[33] δ_H (400 MHz, $CDCl_3$)[33] 1.96, 2.03, 2.04, 2.05, 2.07, 2.09, 2.12, 2.15 (24H, 8 x s, 8 x CH_3CO_2), 3.76 (1H, m, H-5a), 3.86 (2H, m, H-4a, H-5b) 4.04-4.16 (3H, m, H-6a, H-6b, H-6b'), 4.43-4.49 (2H, m, H-1b, H-6a'), 4.94 (1H, dd, $J_{2b,3b}$ 10.2 Hz, $J_{3b,4b}$ 3.2 Hz, H-3b), 5.01-5.13 (2H, m, H-2a, H-2b), 5.24 (1H, at, *J* 9 Hz, H-3a), 5.34 (1H, d, *J* 3.5 Hz, H-4b), 5.66 (1H, d, $J_{1a,2a}$ 8.3 Hz, H-1a); HRMS (ESI) Calcd. For $C_{28}H_{38}O_{19}$ (MNa^+) 701.1905. Found 701.1909.

1-(2'-Propargyloxy)-2,3,5,6,2,3,4,6-hepta-*O*-acetyl-β-D-lactoside 5.36[34]



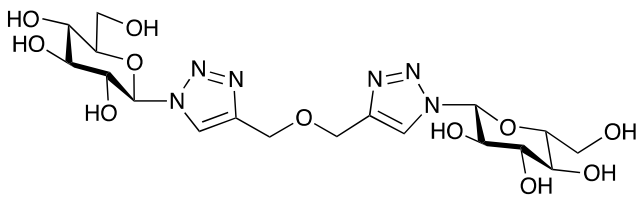
1,2,3,6,2',3',4',6'-Octa-*O*-acetyl-β-D-lactoside **5.35** (24 g, 35.4 mmol) and propargyl alcohol (4.2 mL, 70.7 mmol) were stirred in dry DCM (300 mL) under an N_2 atmosphere and the solution was cooled to 0 °C. Boron trifluoride diethyl etherate (5.3 mL, 42.4 mmol) was added and the mixture was allowed to warm to room temperature. After 16 h, the reaction mixture was diluted with DCM (300 mL), washed with saturated aqueous $NaHCO_3$ (2 x 500 mL), brine (500 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo.* Purification by flash column chromatography (petrol:ethyl acetate 3:2) afforded 1-(2'-propargyloxy)-2,3,5,6,2,3,4,6-hepta-*O*-acetyl-β-D-lactoside **5.36** (10.6 g, 44 %) as a white crystalline solid; m.p. 155-160 °C [lit. 153-155 °C];[35] $[\alpha]_D^{20}$ -25 (*c*, 1.0 in $CHCl_3$) [lit. $[\alpha]_D^{20}$ –23.8 (*c*, 1.0 in $CHCl_3$)];[35] δ_H (400 MHz, $CDCl_3$)[35] 1.96, 2.04, 2.05, 2.06, 2.12, 2.15 (21H, 6 x s, 7 x CH_3CO_2), 2.44 (1H, t, *J* 2.3 Hz, CH_2CH), 3.63 (1H, m, H-5a), 3.78-3.88 (2H, m H-4a, H-5b), 4.07-4.14 (3H, m, H-6a, H-6b, H-6b'), 4.32 (2H, d, *J* 2.4 Hz, CH_2CH), 4.46-4.51 (2H, m, H-1b, H-6a'), 4.73 (1H, d, $J_{1a,2a}$ 7.8 Hz, H-1a), 4.88-4.97 (2H, m, H-2a, H-3b), 5.10 (1H, dd, $J_{2b,3b}$ 10.2 Hz, H-2a), 5.22 (1H, at, *J* 9.0 Hz, H-3a) 5.33 (1H, d, *J* 3.5 Hz, H-4b); HRMS (ESI) Calcd. For $C_{29}H_{38}O_{18}$ (MNa^+) 697.1956. Found 197.1954.

2'-Propargyloxy-β-D-lactoside 5.37[34]



1-(2'-Propargyloxy)-2,3,5,6,2,3,4,6-hepta-*O*-acetyl-β-D-lactoside **5.36** (10 g, 16.7 mmol) was stirred in dry methanol (300 mL), and sodium metal (660 mg) was added portionwise. After 15 h, t.l.c. (ethyl acetate) indicated complete consumption of starting material (R_f 0.7) and the formation of a single product (R_f 0). The reaction mixture was neutralized with Amberlite® IR120 (H^+), filtered, and concentrated *in vacuo* to give 2'-propargyloxy-β-D-lactoside **5.37** (5.7 g, quant.) as a white powder, m.p. 100-105 °C; $[\alpha]_D^{20}$ -39 (*c*, 1.0 in MeOH) δ_H (400 MHz, D_2O)[34] 2.81 (1H, m, $CH_2CC\underline{H}$), 3.23 (1H, at, *J* 7.8 Hz, H-2a), 3.43 (1H, t, *J* 9.0 Hz, H-2b), 3.51-3.58 (4H, m, H-3a, H-3b, H-5a, H-5b), 3.60-3.72 (4H, m, H-4a, H-6a, H-6a', H-6b), 3.81-3.89 (2H, m, H-4b, H-6b'), 4.34 (1H, d, $J_{1b,2b}$ 8.6 Hz, H-1b), 4.36 (2H, bs, $C\underline{H_2}CCH$), 4.56 (1H, d, $J_{1a,2a}$ 7.9 Hz, H-1a); HRMS (ESI) Calcd. For $C_{15}H_{24}O_{11}$ (MNa^+) 403.1216. Found 403.1216.

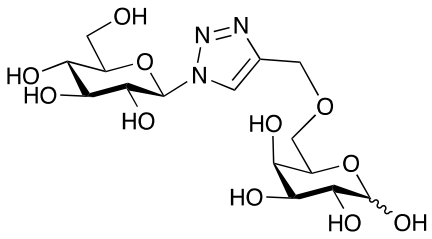
Bis(1-(β-D-glucopyranosyl)-1,2,3-triazol-4-methyl)ether 5.38



D-Glucose **5.15** (500 mg, 2.78 mmol) and triethylamine (1.9 mL, 13.9 mmol) were stirred in D_2O/MeCN (4:1, 10 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (2.4 g, 8.33 mmol) was added. After 1 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.3) and the formation of a major product (R_f 0.5). Propargyl ether (86 μL, 0.834 mmol), $CuSO_4.5H_2O$ (10 mg, 0.04 mmol), and L-ascorbic acid (100 mg, 0.56 mmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, t.l.c. (H_2O/iPrOH/EtOAc, 1:6:3) indicated complete consumption of starting material (R_f 0.8) and the formation of a single product (R_f 0.2). The reaction mixture was concentrated, and the residue was dissolved in ethanol (20 mL), filtered through Celite®, and concentrated *in vacuo*. The residue was then redissolved in water (10 mL), washed with DCM (2 x 20 mL), filtered through a column of Amberlite® IR120 (H^+, previously treated with 1 M NaOH solution, and concentrated *in vacuo*. Purification by flash column chromatography ($CHCl_3$:MeOH, 3:1 → H_2O:iPrOH:EtOAc, 1:6:3) gave bis(1-(β-D-glucopyranosyl)-1,2,3-triazol-4-methyl)ether **5.38** (350 mg, 84 %) as a white solid; m.p. 125-130 °C; $[\alpha]_D^{20}$ -4 (*c*, 1.0 in MeOH); ν_{max} (neat) 3271 (bs, OH), 1030 (s, C-O-C) cm^{-1}; δ_H (500 MHz, D_2O) 3.53 (2H, at *J* 9.4 Hz, 2 x H-4), 3.56 (2H, m, 2 x H-3), 3.58 (2H, m, 2 x H-5), 3.62 (2H, m, 2 x H-6), 3.78 (2H, m, 2 x H-6') 3.89 (2H, t, *J* 9.3 Hz, 2 x H-2) 4.66 (4H, s, 2 x CH_2), 5.64 (2H, d, $J_{1,2}$ 9.0 Hz, 2 x H-1), 8.16

(2H, s, 2 x C=CH); δ_C (125 MHz, D_2O) 60.3 (t, 2 x C-6), 62.2 (t, 2 x CH_2), 68.8 (d, 2 x C-4) 72.1 (d, 2 x C-2), 75.7 (d, 2 x C-3), 78.8 (d, 2 x C-5), 87.3 (d, 2 x C-1), 124.4 (d, 2 x C=<u>C</u>H), 143.7 (s, 2 x <u>C</u>=CH); HRMS (ESI) Calcd. For $C_{18}H_{27}D_2N_6O_{11}$ (MH^+) 507.2020. Found 507.2020.

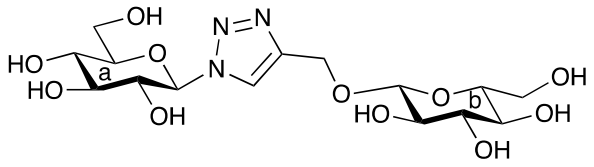
Methyl 6-*O*-(1'-(β-D-glucopyranosyl)-[1',2',3']-triazoly-4'-yl)-α/β-D-galactopyranose 5.39



D-Glucose **5.15** (500 mg, 2.78 mmol) and triethylamine (1.9 mL, 13.9 mmol) were stirred in D_2O/MeCN (4:1, 10 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (2.4 g, 8.33 mmol) was added. After 1 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.3) and the formation of a major product (R_f 0.5). 6-*O*-Prop-2-yn-1-yl-D-galactopyranose **5.30** (1.2 g, 5.56 mmol), $CuSO_4.5H_2O$ (10 mg, 0.04 mmol), and L-ascorbic acid (100 mg, 0.56 mmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.5) and the formation of a single product (R_f 0.1). The reaction mixture was then concentrated, and the residue was dissolved in methanol (20 mL), filtered through Celite®, and concentrated *in vacuo*. The residue was then redissolved in water (10 mL), washed with DCM (2 x 20 mL), filtered through a column of Amberlite® IR120 (H^+, previously treated with 1 M NaOH solution), and

concentrated *in vacuo*. Purification by flash column chromatography ($CHCl_3$:MeOH, 3:1 → H_2O:*i*PrOH:EtOAc, 1:6:3) gave methyl 6-*O*-[1'(β-D-glucopyranosyl)-[1',2',3']-triazoly-4'-yl]-α/β-D-galactopyranose **5.39** (910 mg, 80 %) as a white solid; m.p. 40-45 °C; ν_{max} (neat) 3258 (s, OH), 1018 (s, C-O-C) cm^{-1}; δ_H (500 MHz, D_2O) 3.29 (1H, at, *J* 7.9 Hz, H-$2_{\beta Gal}$), 3.45-3.79 (14H, m, H-4_{Glc}, H-5_{Glc}, H-6_{Glc}, H-6'$_{Glc}$, H$2_{Gal\alpha}$, H-$3_{Gal\alpha}$, H-$3_{Gal\beta}$, H-$4_{Gal\beta}$, H-$5_{Gal\alpha}$, H-$5_{Gal\beta}$, H-$6_{Gal\alpha}$, H-$6_{Gal\beta}$, H-6'$_{Gal\alpha}$, H-6'$_{Gal\beta}$), 3.85 (1H, at, *J* 9.3 Hz, H-2_{Glc}), 4.06 (1H, at, *J* 6.6 Hz, H-$4_{Gal\alpha}$), 4.40 (1H, $J_{1,2}$ 7.8 Hz, H-$1_{Gal\beta}$), 4.60 (2H, s, CH_2), 5.08 (1H, d, $J_{1,2}$ 3.4 Hz, H-$1_{Gal\alpha}$), 5.61 (1H, d, $J_{1,2}$ 9.3 Hz, H-1_{Glc}), 8.13 (1H, s, C=CH); δ_C (125 MHz, D_2O) 60.3 (t, C-6_{Glc}), 63.2 (t, CH_2), 68.1 (d, C-$2_{Gal\alpha}$), 68.6 (d, C-4_{Glc}), 68.8, 68.9 (2 x d, C-4 $_{Gal\alpha\beta}$), 69.3 (t, C-2, C-6 $_{Gal\alpha\beta)}$, 69.4 (t, C-6 $_{Gal\alpha\beta}$), 69.6 (d, C-$2_{Gal\alpha)}$, 71.6 (d, C-2 $_{Gal\beta}$), 72.1 (d, C-2_{Glc}), 72.6, 73.2 (d, C-$5_{Gal\alpha}$, C-$5_{Gal\beta}$), 75.8 (d, C-3_{Glc}), 78.8 (d, C-5_{Glc}), 87.3 (d, C-1_{Glc}), 92.2 (d, C-$1_{Gal\alpha}$), 96.3 (d, C-$1_{Gal\beta}$), 124.4 (d, C=CH), 144.0 (s, C=CH) HRMS (ESI) Calcd. For $C_{15}H_{25}N_3O_{11}$ (MH^+) 424.1567. Found 424.1559.

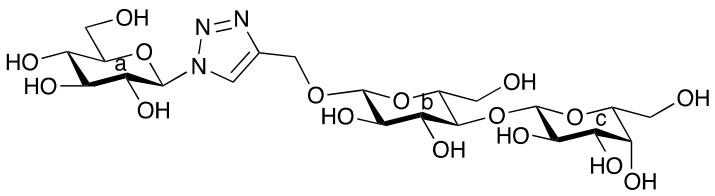
Methyl (1'-(β-D-glucopyranosyl)-[1',2',3']-triazoly-4'-yl)-β-D-glucopyranoside 5.40



D-Glucose **5.4** (500 mg, 2.78 mmol) and triethylamine (1.9 mL, 13.9 mmol) were stirred in D_2O/MeCN (4:1, 10 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (2.4 g, 8.33 mmol) was added. After

1 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.3) and the formation of a major product (R_f 0.5). Prop-2-yn-1-yl β-D-glucopyranoside **5.33** (1.2 g, 5.56 mmol), $CuSO_4.5H_2O$ (10 mg, 0.04 mmol), and L-ascorbic acid (100 mg, 0.56 mmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, t.l.c. (H_2O/iPrOH/EtOAc, 1:6:3) indicated complete consumption of starting material (R_f 0.8) and formation of a single product (R_f 0.2). The reaction mixture was concentrated, and the residue was dissolved in ethanol (20 mL), filtered through Celite®, and concentrated *in vacuo*. The residue was then redissolved in water, washed with DCM (2 x 20 mL), filtered through a column of Amberlite® IR120 (H^+, previously treated with 1 M NaOH solution), and concentrated *in vacuo*. Purification by flash column chromatography ($CHCl_3$:MeOH, 3:1 → H_2O:iPrOH:EtOAc, 1:6:3) gave methyl (1'-(β-D-glucopyranosyl)-[1',2',3']-triazoly-4'-yl)-β-D-glucopyranoside **5.40** (950 mg, 84 %) as a white solid; m.p. 65-70 °C; $[\alpha]_D^{20}$ -21 (*c*, 1.0 in MeOH); ν_{max} (neat) 3273 (s, OH), 1030 (s, O-C-O) cm^{-1}; δ_H (500 MHz, D_2O) 3.12 (1H, at, *J* 8.6 Hz, H-2b), 3.22 (1H, at, *J* 9.5 Hz, H-3b), 3.29-3.33 (2H, m, H-4a, H-4b), 3.44-3.64 (5H, m, H-3a, H-5a, H-5b, H-6a, H-6b), 3.75 (2H, m, H-6a', H-6b'), 3.84 (1H, at, *J* 12.7 Hz, H-2a), 4.40 (1H, d, $J_{1b,2b}$ 7.9 Hz, H-1b), 4.80 (2H, m, CH_2), 5.61 (1H, d, $J_{1a,2a}$ 9.0 Hz, H-1a), 8.14 (1H, s, C=CH); δ_C (125 MHz, D_2O) 60.3 (t, C-6a), 60.8 (t, C-6b), 61.8 (t, CH_2), 68.6 (d, C-4a), 69.4 (d, C-4b), 72.1 (d, C-2a), 72.9 (d, C-2b), 75.6 (d, C-5a), 75.8 (d, C-5b), 75.8 (d, C-3a), 78.8 (d, C-3b), 87.3 (d, C-1a), 101.4 (d, C-1b) 124.6 (d, CH=C), 143.7 (s, CH=C); HRMS (ESI) Calcd. For $C_{15}H_{26}N_3O_{11}$ (MH^+) 424.1567. Found 424.1563.

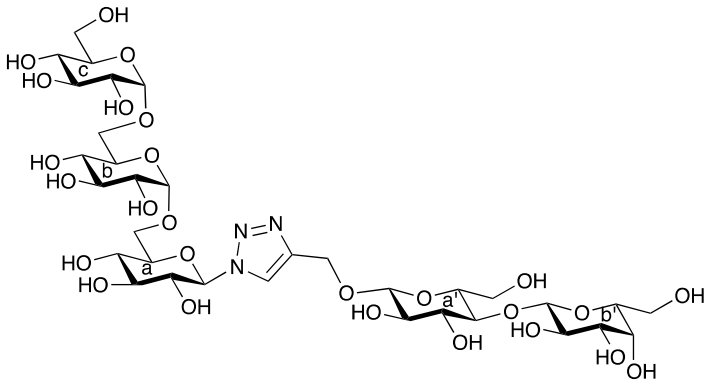
Methyl (1'(β-D-glucopyranosyl)-[1',2',3']-triazoly-4'-yl) β-D-lactopyranoside 5.41



D-Glucose **5.4** (500 mg, 2.78 mmol) and triethylamine (1.9 mL, 13.9 mmol) were stirred in D_2O/MeCN (4:1, 10 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (2.4 g, 8.33 mmol) was added. After 1 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.3) and the formation of a major product (R_f 0.5). 2'-Propargyloxy-β-lactoside **5.37** (840 mg, 2.21 mmol), $CuSO_4.5H_2O$ (10 mg, 0.04 mmol), and L-ascorbic acid (100 mg, 0.56 mmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, t.l.c. ($CHCl_3$:MeOH, 2:1) indicated complete consumption of starting material (R_f 0.5) and formation of a single product (R_f 0). The reaction mixture was filtered through a column of Amberlite® IR120 (H^+, previously treated with 1 M NaOH solution, and concentrated *in vacuo*. The reaction mixture was purified by gel filtration on a column (8.7 x 1.7 cm) of Sephadex G-25 (pre-equilibrated and eluted with 0.01% ammonia) to give methyl (1'(β-D-glucopyranosyl)-[1',2',3']-triazoly-4'-yl)-β-D-lactopyranoside **5.41** (990 mg, 75 %) as a white solid; m.p. 100-105 °C; $[\alpha]_D^{20}$ -4 (*c*, 1.0 in H_2O); ν_{max} (neat) 3270 (s, OH), 1031 (s, O-C-O) cm^{-1}; δ_H (500 MHz, D_2O) 3.18 (1H, at, *J* 8.3 Hz, H-2c), 3.38 (1H, at, *J* 7.8 Hz, H-2b), 3.48-3.50 (5H, m, H-3b, H-3c, H-4b, H-4c, H-5a, H-5b), 3.57 (2H, m, H-3a, H-5c), 3.61-3.66 (4H, m, H-6a, H-6b, H-6'b, H-6c), 3.77 (1H, m, H-6'a), 3.81 (1H, m, H-6'c), 3.86 (1H, at, *J* 9.2 Hz, H-2a), 4.29 (1H, d, $J_{1c,2c}$ 7.9 Hz, H-1c), 4.44 (1H, d, $J_{1b,2b}$ 8.1 Hz, H-1b), 4.77-4.89 (2H, m,

CH_2), 5.61 (1H, d, $J_{1a,2a}$ 9.0 Hz, H-1a), 8.16 (1H, s, C=CH); δ_C (125 MHz, D_2O) 59.9 (t, C-6c), 60.3 (t, C-6a), 60.9 (t, C-6b), 61.9 (t, CH_2), 68.4 (d, C-4c), 68.8 (d, C-4b), 70.9 (d, C-2c), 72.2 (d, C-2a), 72.4 (d, C-4a), 72.6 (d, C-2b), 74.2 (d, C-5a), 74.7 (d, C-5b), 75.3 (d, C-5c), 75.8 (d, C-3a), 78.1 (d, C-3b), 78.8 (d, C-3c), 87.3 (d, C-1a), 101.3 (d, C-1b), 102.8 (d, C-1c), 124.5 (d, CH=C), 143.9 (s, CH=C); HRMS (ESI) Calcd. For $C_{21}H_{35}N_3O_{16}$ (MNa^+) 608.1915. Found 608.1907.

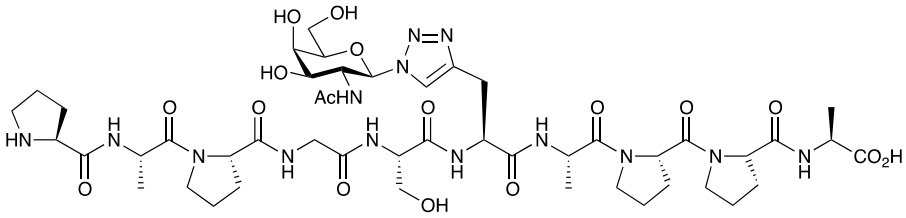
Methyl (1'-(α-D-glucosyl-[1→6]-α-D-glucosyl-[1→6]-β-D-glucopyranosyl)-[1',2',3']-triazoly-4'-yl)-β-D-lactopyranoside 5.42



Isomaltotriose **5.21** (100 mg, 0.20 mmol) and triethylamine (0.14 mL, 0.99 mmol) were stirred in D_2O/MeCN (4:1, 2 mL) and cooled to 0 °C. 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (170 mg, 0.60 mmol) was added. After 1 h, 2'-propargyloxy-β-lactoside **5.37** (150 mg, 0.40 mmol), $CuSO_4.5H_2O$ (1 mg, 3 μmol), and L-ascorbic acid (7 mg, 0.04 mmol) were added and the reaction mixture was stirred at 50 °C overnight. The reaction mixture was then filtered through a column of Amberlite® IR120 (H^+, previously treated with 1 M NaOH solution, and concentrated *in vacuo*. The reaction mixture was purified by gel filtration on a column (8.7 x 1.7 cm) of Sephadex G-25 (pre-equilibrated and eluted with 0.01% ammonia) to give methyl (1'-(α-D-glucosyl-[1→6]-α-D-glucosyl-[1→6]-β-D-glucopyranosyl)-[1',2',3']-triazoly-4'-yl)-β-D-lactopyranoside **5.42** (145 mg, 81 %) as a white solid, m.p. 180-185 °C; $[\alpha]_D^{20}$ +52 (*c*, 1.0 in H_2O); ν_{max} (neat) 3313 (bs, OH), 1019 (s, O-C-O) cm^{-1}; δ_H (400 MHz, D_2O) 3.19 (1H, at, *J* 8.6 Hz, H-2b'), 3.25-4.43 (4H, m, H-2b, H-4b, H-2a', H-3a'), 3.46-3.69 (18H, m, H-3a, H-4a, H-6a, H-6'a, H-3b, H-5b, H-6', H-6'b, H-2c, H-3c, H-4c, H-5c, H-4a', H-5a', H-6a', H-6'a', H-3b', H-5'b), 3.76-3.82 (5H, m, H-5a, H-6c, H-6'c, H-6b', H-6'b'), 3.92 (1H, t, *J* 9.1 Hz, H-2a), 4.29 (1H, d, $J_{1b',2b'}$ 9.1 Hz,

H-1b'), 4.46 (1H, d, $J_{1a'2a'}$ 8.1 Hz, H-1a'), 4.71 (1H, d, $J_{1c,2c}$ 3.7 Hz, H-1c), 4.77-4.90 (3H, m, H-1b, CH_2), 5.64 (1H, d, $J_{1a,2a}$ 9.0 Hz, H-1a), 8.21 (1H, s, C=CH); δ_C (100 MHz, D_2O) 60.0 (t, C-6b'), 60.4 (t, C-6a), 61.0 (t, C-6a'), 62.0 (t, CH_2), 65.0 (d, C-6b) 66.3 (t, C-6c), 68.5 (d, C-4b'), 69.1 (d, C-4c), 69.2 (d, C-4a'), 69.4 (d, C-4b), 70.1 (d, C-2c), 70.9 (d, C-2b'), 71.2 (d, C-2b), 71.4 (d, C-3c), 71.7 (d, C-3b), 72.0 (d, C-2a), 72.5 (d, C-4a), 72.7 (d, C-2a'), 73.1 (d, C-5b), 73.2 (d, C-5c), 74.3 (d, C-3a'), 74.8 (d, C-5a'), 75.3 (d, C-5b'), 76.1 (d, C-3a), 77.5 (d, C-5a), 78.2 (d, C-3b'), 87.3 (d, C-1a), 97.7 (d, C-1c), 98.1 (d, C-1b), 101.4 (d, C-1a), 102.9 (d, C-1b'), 124.8 (d, CH=C) 143.9 (s, CH=C); HRMS (ESI) Calcd. For $C_{33}H_{55}N_3O_{26}$ (MNa^+) 932.2971. Found 932.2967.

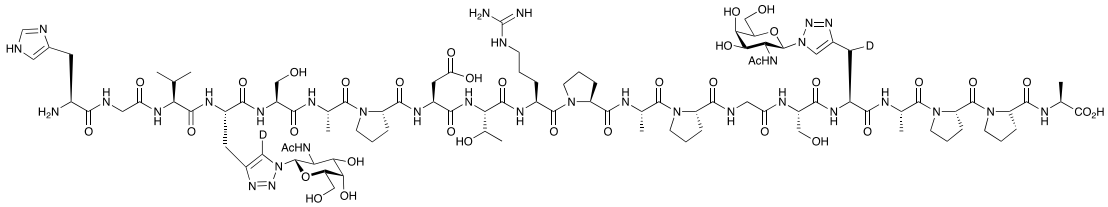
Click Neoglycopeptide **5.45**



N-Acetyl-D-galactosamine **5.13** (0.52 mg, 2.33 μmol) and triethylamine (1.6 μL, 11.6 μmol) were dissolved in D_2O (7.46 μL). 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (2 mg, 6.99 μmol) dissolved in MeCN (1.86 μL) was added and the reaction mixture was incubated at rt for 6 h, at which time, 1.2 M HCl (10 μL) was added until the solution was acidic (pH 2), followed by sat. $NaHCO_3$ (9.5 μL). Then, peptide **5.43** (1 mg, 1.16 μmol) dissolved in water (10 μL), $CuSO_4.5H_2O$ (250 μg, 1 μmol), and L-ascorbic acid (180 μg, 1 μmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, the crude product was isolated by analytical RP HPLC (column: Jupiter 5U C18 (300 Å) column (Phenomenex); eluent:

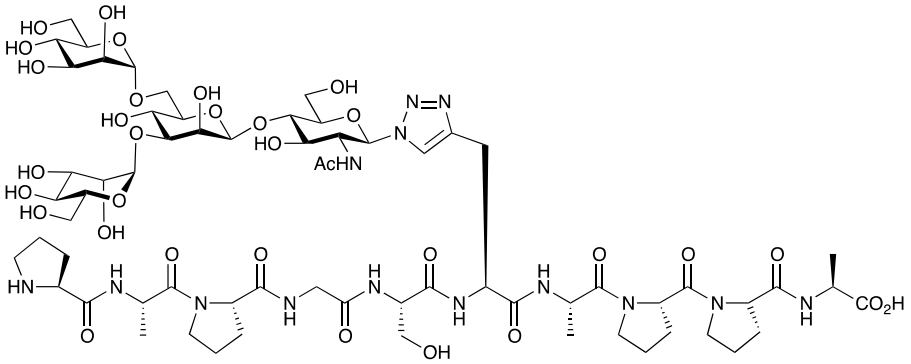
A (0.05% TFA in H_2O) and B (0.1% TFA in MeCN); gradient: sample was run at 1 mL/min with a gradient of 5-65% B over 40 mins; column oven: 40 °C; detection: UV 235 nm) to afford compound **5.45** (0.6 mg, 47 %) as a white solid; HRMS (ESI) Calcd. For $C_{47}H_{72}N_{14}O_{17}$ (MH^+) 1105.5278. Found 1105.5253.

Click Neoglycopeptide **5.46**



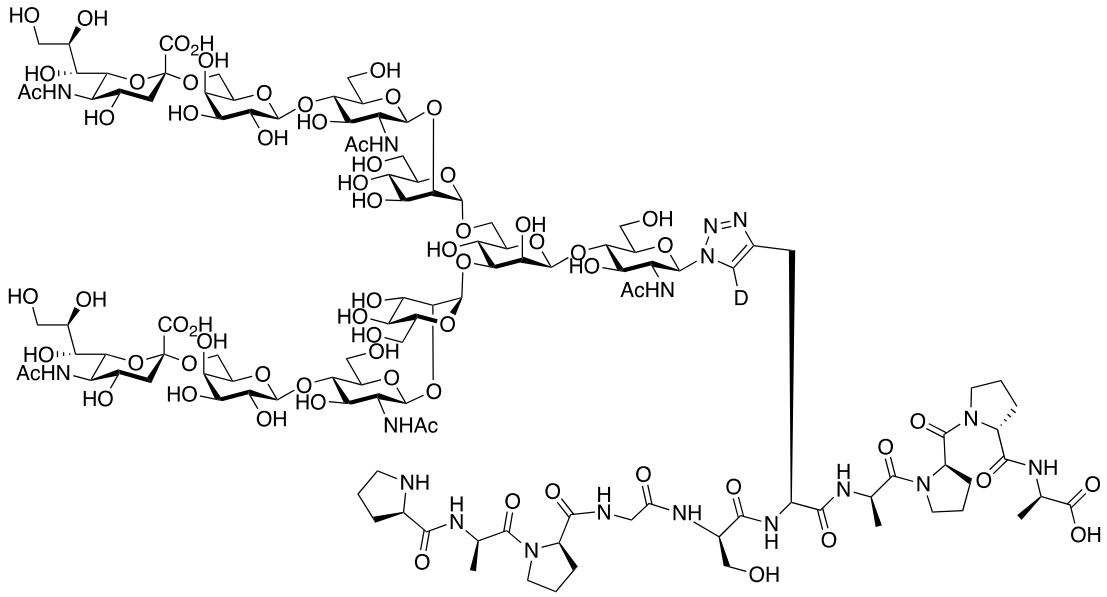
N-Acetyl-D-galactosamine **5.13** (0.47 mg, 2.13 μmol) and triethylamine (1.48 μL, 10.7 μmol) were dissolved in D_2O (6.8 μL). 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (1.8 mg, 6.4 μmol) dissolved in MeCN (1.7 μL) was added and the reaction mixture was incubated at rt for 6 h, at which time, 1.2 M HCl (7 μL) was added until the solution was acidic (pH 2), followed by sat. $NaHCO_3$ (13.4 μL). Then, peptide **5.44** (1 mg, 0.53 μmol) dissolved in water (10 μL), $CuSO_4.5H_2O$ (250 μg, 1 μmol), and L-ascorbic acid (180 μg, 1 μmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, the crude product was isolated by analytical RP HPLC (column: Jupiter 5U C18 (300 Å) column (Phenomenex); eluent: A (0.05% TFA in H_2O) and B (0.1% TFA in MeCN); gradient: sample was run at 1 mL/min with a gradient of 5-65% B over 40 mins; column oven: 40 °C; detection: UV 235 nm) to afford compound **5.46** (600 μg, 46%) as a white solid; HRMS (ESI) Calcd. For $C_{98}H_{151}N_{33}O_{36}$ (MH^+) 2367.1078. Found 2367.1023.

Click Neoglycopeptide **5.48**



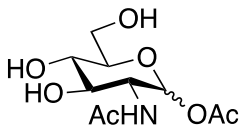
Tetrasaccharide **5.47** (1.6 mg, 2.33 μmol) and triethylamine (1.6 μL, 11.6 μmol) were dissolved in D_2O (7.46 μL). 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (2 mg, 6.99 μmol) dissolved in MeCN (1.86 μL) was added and the reaction mixture was incubated at rt for 6 h, at which time, 1.2 M HCl (10 μL) was added until the solution was acidic (pH 2), followed by sat. $NaHCO_3$ (9.5 μL). Then, peptide **5.43** (1 mg, 1.16 μmol) dissolved in water (10 μL), $CuSO_4.5H_2O$ (250 μg, 1 μmol), and L-ascorbic acid (180 μg, 1 μmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, the crude product was isolated by analytical RP HPLC (column: Jupiter 5U C18 (300 Å) column (Phenomenex); eluent: A (0.05% TFA in H_2O) and B (0.1% TFA in MeCN); gradient: sample was run at 1 mL/min with a gradient of 5-65% B over 40 mins; column oven: 40 °C; detection: UV 235 nm) to afford compound **5.48** (0.5 mg, 27 %) as a white solid; HRMS (ESI) Calcd. For $C_{65}H_{101}DN_{14}O_{32}$ (MH^+) 1591.6847. Found 1591.6838.

Click Neoglycopeptide **5.50**



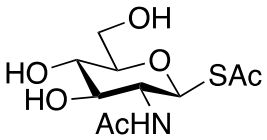
Sialoglycan **5.49** (4.7 mg, 2.33 μmol) and triethylamine (1.6 μL, 11.6 μmol) were dissolved in D_2O (7.46 μL). 2-Azido-1,3-dimethylimidazolinium hexafluorophosphate **5.10** (2 mg, 6.99 μmol) dissolved in MeCN (1.86 μL) was added and the reaction mixture was incubated at rt for 6 h, at which time, 1.2 M HCl (10 μL) was added until the solution was acidic (pH 2), followed by sat. $NaHCO_3$ (9.5 μL). Then, peptide **5.43** (1 mg, 1.16 μmol) dissolved in water (10 μL), $CuSO_4.5H_2O$ (250 μg, 1 μmol), and L-ascorbic acid (180 μg, 1 μmol) were added and the reaction mixture was stirred at 50 °C for 14 h, after which time, the crude product was isolated by analytical RP HPLC (column: Jupiter 5U C18 (300 Å) column (Phenomenex); eluent: A (0.05% TFA in H_2O) and B (0.1% TFA in MeCN); gradient: sample was run at 1 mL/min with a gradient of 5-65% B over 40 mins; column oven: 40 °C; detection: UV 235 nm) to afford compound **5.50** (0.6 mg, 42 %) as a white solid; HRMS (ESI) Calcd. For $C_{115}H_{180}DN_{18}O_{68}$ (MH^+) 2904.1400. Found 2904.1376.

1-Acetyl-2-Acetamido-2-deoxy-D-glucopyranoside 5.52



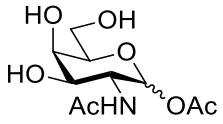
N-Acetyl-D-glucosamine **5.4** (500 mg, 2.26 mmol), thioacetic acid (0.8 mL, 11.3 mmol), and triethylamine (3.1 mL, 22.6 mmol) were stirred in H_2O (10 mL) and cooled to 0 °C. 2-Chloro-1,3-dimethylimidazolinium chloride **5.1** (1.2 g, 6.78 mmol) was added. After 30 mins, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.2) and the formation of a major product (R_f 0.5). The reaction mixture was then diluted with water (20 mL), washed with DCM (5 x 20 mL), concentrated to 1 mL, filtered through a column of Amberlite® IR120 (H^+) previously activated by 1 M NaOH solution, and concentrated *in vacuo*. Purification by flash column chromatography ($CHCl_3$:MeOH, 5:1) gave 1-acetyl-2-acetamido-2-deoxy-D-glucopyranoside **5.52** (510 mg, 86 %) as a white solid (α/β, 3:2); δ_H (400 MHz, D_2O) 1.83 (3H, s, α-$NHCOCH_3$), 1.89 (3H, s, β-$NHCOCH_3$), 3.36-3.55 (4H, m, H-4α, H-3β, H-4β, H-5β), 3.61-3.80 (6H, m, H-3α, H-5α, H-2β, H-6α, H-6'α, H-6β, H-6'β), 3.93 (1H, dd, $J_{1\alpha,2\alpha}$ 3.6 Hz, $J_{2\alpha,3\alpha}$ 10.6 Hz, H-2α), 5.51 (1H, d, $J_{1\beta,2\beta}$ 8.7 Hz, H-1), 5.95 (1H, d, H-1α); δ_C (100 MHz, D_2O) 19.9, 20.1 (2 x q, 2 x $NHCOCH_3$), 21.6 (q, β-$CO_2\underline{C}H_3$), 21.9 (q, α-$CO_2\underline{C}H_3$), 52.4 (d, C-2α), 54.4 (d, C-2β), 60.1, 60.2 (2 x t, 2 x C-6), 69.2 (d, C-4β), 69.3 (d, C-4α), 70.4 (d, C-3α), 73.2 (d, C-3β), 73.9 (d, C-5α), 76.6 (d, C-5β), 90.9 (d, C-1α), 92.8 (d, C-1β), 172.1, 172.7, 174.6, 174.8 (4 x s, 4 x C=O); HRMS (ESI) Calcd. For $C_{10}H_{17}NO_7$ (MNa^+) 286.0903. Found 286.0904.

1-Thioacetyl-2-Acetamido-2-deoxy-β-D-glucopyranoside 5.54



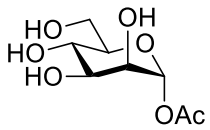
N-Acetyl-D-glucosamine **5.4** (500 mg, 2.26 mmol) and triethylamine (3.1 mL, 22.6 mmol) were stirred in D_2O (10 mL) and cooled to 0 °C. 2-Chloro-1,3-dimethylimidazolinium chloride **5.1** (1.2 g, 6.78 mmol) was added. After 30 mins, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.2) and the formation of a major product (R_f 0.6). Thioacetic acid (2.4 mL, 33.9 mmol) was added dropwise and the reaction mixture was allowed to stir for an additional 30 mins at 0 °C. The reaction mixture was then diluted with water (10 mL) and washed with DCM (5 x 20 mL). The aqueous phase was concentrated to 1 mL, filtered through a column of Amberlite® IR120 (H^+, previously activated by 1 M NaOH solution), and concentrated *in vacuo*. Purification by flash column chromatography ($CHCl_3$:MeOH, 5:1) gave 1-thioacetyl-2-acetamido-2-deoxy-β-D-glucopyranoside **5.54** (570 mg, 90 %) as a white solid, m.p. 195-200 °C; $[\alpha]_D^{20}$ +24 (*c*, 1.0 in MeOH); δ_H (400 MHz, D_2O) 1.89 (3H, s, $NHCOCH_3$), 2.31 (3H, s, $SCOCH_3$), 3.40 (1H, at, *J* 9.0 Hz, H-4), 3.46 (1H, m, H-5), 3.52 (1H, at, *J* 9.0 Hz, H-3), 3.62 (1H, dd, $J_{5,6}$ 5.0 Hz, $J_{6,6'}$ 12.3 Hz, H-6), 3.75-3.82 (2H, m, H-2, H-6'), 5.10 (1H, d, $J_{1,2}$ 10.6 Hz, H-1); δ_C (100 MHz, D_2O) 22.0 (q, $NHCO\underline{C}H_3$), 30.2 (q, $SCO\underline{C}H_3$), 53.4 (d, C-2), 60.5 (t, C-6), 69.4 (d, C-4), 74.9 (d, C-3), 80.4 (d, C-5), 80.9 (d, C-1), 174.5 (s, $NH\underline{C}OCH_3$), 197.6 (s, $S\underline{C}OCH_3$); HRMS (ESI) Calcd. For $C_{10}H_{18}NO_6S$ (MH^+) 280.0855. Found 280.0848.

Acetyl 2-Acetamido-2-deoxy-D-galactopyranoside 5.55



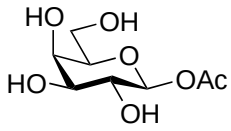
N-Acetyl-D-galactosamine **5.13** (20 mg, 0.09 mmol), triethylamine (125 μL, 0.9 mmol), and thioacetic acid (32 μL, 0.452 mmol) were stirred in D_2O (500 μL) and cooled to 0 °C. 2-Chloro-1,3-dimethylimidazolinium chloride **5.1** (46 mg, 0.27 mmol) was added. After 30 mins, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.2) and the formation of a major product (R_f 0.5). The reaction mixture was diluted with water (10 mL), washed with DCM (5 x 20 mL), lyophilised to 1 mL, filtered through a column of Amberlite® IR120 (H^+) previously activated by 1 M NaOH solution, and then lyophilized. The product was purified by semi-preparative RP HPLC (column: Luna 5U C18 (100 Å) column (Phenomenex); eluent: H_2O only. Sample was run at 1 mL/min for 40 mins; column oven: 40 °C; detection: CAD) to afford gave acetyl 2-acetamido-2-deoxy-D-galactopyranoside **5.55** (18 mg, 86 %) as a white foam (α/β, 1:1.1), t_R = 11.65 mins; ν_{max} (neat) 3273 (s, OH), 1742 (s, C=O), 1642 (s, N-H) cm^{-1}; δ_H (400 MHz, D_2O) 1.91 (6H, s, α-NHCOCH_3, β-NHCOCH_3), 2.04 (3H, s, β-COCH_3), 2.08 (3H, s, α-CO<u>C</u>H_3), 3.63-3.68 (7H, H-4β, H-5α, H-5β, H-6α, H-6'α, H-6β, H-6'β), 3.89-3.97 (4H, m, H-2α, H-3α, H-3β, H-4α), 4.20 (1H, m, H-2α), 5.49 (1H, d, $J_{1,2}$ 8.6 Hz, H-1β), 6.01 (1H, s, H-1α); δ_C (100 MHz, D_2O) 20.2, 20.2, 21.7, 22.0 (4 x q, 4 x CH_3CO_2), 48.7 (d, C-2α), 51.3 (d, C-2β), 60.7, 61.0 (2 x t, C-6α. C-6β), 67.2, 67.5 (2 x d, C-3α, C-3β), 68.2, 70.4 (2 x d, C-4α, C-4β), 73.2, 76.1 (2 x d, C-5α, C-5β), 91.2 (d, C-1α), 93.3 (d, C-1β), 172.4, 172.8, 174.8, 175.0 (4 x s, 4 x <u>C</u>OCH_3); HRMS (ESI) Calcd. For $C_{10}H_{17}NO_7$ (MNa^+) 286.0903. Found 286.0904.

Acetyl α-D-mannopyranoside 5.56



D-Mannose **5.17** (500 mg, 2.78 mmol), triethylamine (3.8 mL, 27.7 mmol), and thioacetic acid (1 mL, 13.8 mmol) were stirred in D_2O (10 mL) and cooled to 0 °C. 2-Chloro-1,3-dimethylimidazolinium chloride **5.1** (1.4 g, 8.33 mmol) was added. After 2 h, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.3) and the formation of a major product (R_f 0.6). The reaction mixture was then diluted with water (90 mL), washed with DCM (3 x 100 mL), lyophilised to 1 mL, filtered through a column of Amberlite® IR120 (H^+) previously activated by 1 M NaOH solution, and then lyophilized. The product was purified by flash column chromatography (10:1 $CHCl_3$:MeOH → 5:1 $CHCl_3$:MeOH**)** to afford acetyl α-D-galactopyranoside **5.56** (450 mg, 73 %) as a white foam; $[\alpha]_D^{20}$ +72 (*c*, 1.0 in MeOH); ν_{max} (neat) (s, OH), (s, C=O), (s, N-H) cm^{-1}; δ_H (400 MHz, D_2O) 2.04 (3H, s, α-$COCH_3$), 3.59-3.65 (3H, m, H-4, H-5, H-6), 3.71-3.73 (1H, m, H-6'), 3.77 (1H, m, H-3), 3.87 (1H, m, H-2), 5.86 (1H, d, $J_{1,2}$ 1.8 Hz, H-1); δ_C (100 MHz, D_2O) 20.2 (q, $CO_2\underline{C}H_3$), 60.6 (t, C-6), 66.1 (d, C-5), 68.9 (d, C-2), 69.0 (d, C-3), 70.1 (d, C-3), 74.7 (d, C-4), 93.6 (d, C-1), 172.3 (s, C=O); HRMS (ESI) Calcd. For $C_8H_{14}O_7$ (MNa^+) 245.0637. Found 245.0594.

Acetyl β-D-galactopyranoside 5.57



D-Galactose **5.57** (20 mg, 0.111 mmol), triethylamine (154 μL, 1.11 mmol), and thioacetic acid (40 μL, 0.555 mmol) were stirred in D_2O (500 μL) and cooled to 0 °C. 2-Chloro-1,3-dimethylimidazolinium chloride **5.1** (57 mg, 0.333 mmol) was added. After 30 mins, t.l.c. ($CHCl_3$/MeOH, 2:1) indicated complete consumption of starting material (R_f 0.3) and the formation of a major product (R_f 0.6). The reaction mixture was then diluted with water (10 mL), washed with DCM (5 x 20 mL), lyophilised to 1 mL, filtered through a column of Amberlite® IR120 (H^+) previously activated by 1 M NaOH solution, and then lyophilized. The product was purified by semi-preparative RP HPLC (column: Luna 5U C18 (100 Å) column (Phenomenex); eluent: H_2O only. Sample was run at 1 mL/min for 40 mins; column oven: 40 °C; detection: CAD) to afford gave acetyl β-D-galactopyranoside **5.58** (21 mg, 85 %) as a white solid, t_R = 6.5 mins; m.p. 160-165 °C; ν_{max} (neat) 3330 (s, OH), 1730 (s, C=O) cm^{-1}; δ_H (400 MHz, D_2O) 2.08 (3H, s, $COCH_3$), 3.59-3.64 (4H, m, H-2, H-3, H-6, H-6'), 3.71 (1H, at, *J* 6.3 Hz, H-5), 4.64 (1H, s, H-4), 5.39 (1H, d, $J_{1,2}$ 7.1 Hz, H-1); δ_C (100 MHz, D_2O) 20.3 (q, CH_3), 60.7 (t, C-6), 68.3 (d, C-4), 69.5 (d, C-2), 72.3 (d, C-3), 76.0 (d, C-5), 94.3 (d, C-1), 172.7 (s, C=O); HRMS (ESI) Calcd. For $C_8H_{14}NO_7$ (MNa^+) 245.0637. Found 245.0641.

References

(1) Thompson, J. D.; Gibson, T. J.; Higgins, D. G. In *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2002.

(2) Sudibya, H. G.; Ma, J.; Dong, X.; Ng, S.; Li, L. J.; Liu, X. W.; Chen, P. *Angew. Chem. Int. Ed.* **2009**, *48* (15), 2723–2726.

(3) Lim, D.; Brimble, M. A.; Kowalczyk, R.; Watson, A. J. A.; Fairbanks, A. J. *Angew. Chem. Int. Ed.* **2014**, *53* (44), 11907–11911.

(4) Wolfrom, M. L.; Thompson, A.; Inatome, M. *J. Am. Chem. Soc.* **1957**, *79* (14), 3868–3871.

(5) Liu, Z.; Byun, H.-S.; Bittman, R. *Org. Lett.* **2010**, *12* (13), 2974–2977.

(6) Mukthavaram, R.; Marepally, S.; Venkata, M. Y.; Vegi, G. N.; Sistla, R.; Chaudhuri, A. *Biomaterials* **2009**, *30* (12), 2369–2384.

(7) Ramiz, M. M. M.; El-Sayed, W. A.; Hagag, E.; Abdel-Rahman, A. A.-H. *J. Heterocycl. Chem.* **2011**, *48* (5), 1028–1038.

(8) Bukowski, R.; Morris, L. M.; Woods, R. J.; Weimar, T. *Eur. J. Org. Chem.* **2001**, 2697–2705.

(9) Zhao, J.; Wei, S.; Ma, X.; Shao, H. *Carbohydr. Res.* **2010**, *345* (1), 168–171.

(10) Czernecki, S.; Randriamandimby, D. *Tetrahedron Lett.* **1993**, *34* (49), 7915–7916.

(11) Czernecki, S.; Ayadi, E. *Can. J. Chem.* **1995**, *73* (3), 343–350.

(12) Chang, S.-S.; Lin, C.-C.; Li, Y.-K.; Mong, K.-K. T. *Carbohydr. Res.* **2009**, *344* (4), 432–438.

(13) Guazzelli, L.; Catelani, G.; D'Andrea, F.; Giannarelli, A. *Carbohydr. Res.* **2009**, *344* (3), 298–303.

(14) Mayer, T. G.; Schmidt, R. R. *Building* **1999**.

(15) Grundler, G.; Schmidt, R. R. *Liebigs Ann. Chem* **1984**, *1984* (11), 1826–1847.

(16) Hollinger, M.; Abraha, F.; Oscarson, S. *Carbohydr. Res.* **2011**, *346* (12), 1454–1466.

(17) Pilgrim, W.; Murphy, P. V. *J. Org. Chem.* **2010**, *75* (20), 6747–6755.

(18) Matta, K. L.; Barlow, J. J. *Carbohydr. Res.* **1976**, *48* (1), 65–71.

(19) Murata, T.; Itoh, T.; Hayakawa, Y.; Usui, T. *J. Biochem.* **1996**, *120* (4), 851–855.

(20) Minuth, T.; Irmak, M.; Groschner, A.; Lehnert, T.; Boysen, M. M. K. *Eur. J. Org. Chem.* **2009**, *2009* (7), 997–1008.

(21) Akiya, S.; Osawa, T. *Chem. Pharm. Bull.* **1960**, *8* (7), 583–587.

(22) Hartman, M. C. T.; Coward, J. K. *J. Am. Chem. Soc.* **2002**, *124* (34), 10036–10053.

(23) Bell, J. R.; Luo, H.; Dai, S. *Tetrahedron Lett.* **2011**, *52* (29), 3723–3725.

(24) Isobe, T.; Ishikawa, T. *J. Org. Chem.* **1999**, *64* (19), 6984–6988.

(25) Raab, V.; Harms, K.; Sundermeyer, J.; Kovačević, B.; Maksić, Z. B. *J. Org. Chem.* **2003**, *68* (23), 8790–8797.

(26) Kitamura, M.; Tashiro, N.; Miyagawa, S.; Okauchi, T. *Synthesis.* **2011**, *7*, 1037–1044.

(27) Carpino, L. A.; El-Faham, A. *J. Am. Chem. Soc.* **1995**, *117* (19), 5401–5402.

(28) Floyd, N.; Vijayakrishnan, B.; Koeppe, J. R.; Davis, B. G. *Angew. Chem. Int. Ed.* **2009**, *48* (42), 7798–7802.

(29) Tanaka, T.; Nagai, H.; Noguchi, M.; Kobayashi, A.; Shoda, S.-I. *Chem. Commun.* **2009**, *23*, 3378–3379.

(30) Ito, H.; Kamachi, T.; Yashima, E. *Chem. Commun.* **2012**, *48* (45), 5650.

(31) Hoheisel, T. N.; Frauenrath, H. *Org. Lett.* **2008**, *10* (20), 4525–4528.

(32) Lu, W.-Y.; Sun, X.-W.; Zhu, C.; Xu, J.-H.; Lin, G.-Q. *Tetrahedron* **2010**, *66* (3), 750–757.

(33) Hronowski, L. J. J.; Szarek, W. A.; Hay, G. W.; Krebs, A.; Depew, W. T. *Carbohydr. Res.* **1989**, *190* (2), 203–218.

(34) Otsuka, A.; Sakurai, K.; Hasegawa, T. *Chem. Commun.* **2009**, 5442.

(35) Altieri, S.; Balzi, M.; Bortolussi, S.; Bruschi, P.; Ciani, L.; Clerici, A. M.; Faraoni, P.; Ferrari, C.; Gadan, M. A.; Panza, L.; Pietrangeli, D.; Ricciardi, G.; Ristori, S. *J. Med. Chem.* **2009**, *52* (23), 7829–7835.